ARMY RESEARCH LABORATORY

Validation of a Task Network Human Performance Model of Driving

by Josephine Q. Wojciechowski

ARL-RP-0174 **April 2007**

A reprint from the thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Master of Science In Industrial and Systems Engineering

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-RP-0174 **April 2007**

Validation of a Task Network Human Performance Model of Driving

Josephine Q. Wojciechowski
Human Research & Engineering Directorate, ARL

A reprint from the thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Master of Science In Industrial and Systems Engineering

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* April 2007	**2. REPORT TYPE** Reprint	**3. DATES COVERED (From - To)** January 2004 to April 2006

4. TITLE AND SUBTITLE Validation of a Task Network Human Performance Model of Driving	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Josephine Q. Wojciechowski (ARL)	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Human Research & Engineering Directorate Aberdeen Proving Ground, MD 21005-5425	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-RP-0174
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBERS

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

A reprint from the thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Master of Science In Industrial and Systems Engineering

14. ABSTRACT

Human performance modeling (HPM) is often used to investigate systems during all phases of development. HPM was used to investigate function allocation in crews for future combat vehicles. The tasks required by the operators centered around three primary functions, commanding, gunning, and driving. In initial investigations, the driver appeared to be the crew member with the highest workload.

Validation of the driver workload model (DWM) is necessary for confidence in the ability of the model to predict workload. Validation would provide mathematical proof that workload of driving is high and that additional tasks impact the performance.

This study consisted of two experiments. The purpose of each experiment was to measure performance and workload while driving and attending to an auditory secondary task. The first experiment was performed with a human performance model. The second experiment replicated the same conditions in a human-in-the-loop driving simulator. The results of the two experiments were then correlated to determine if the model could predict performance and workload changes.

The results of the investigation indicate that there is some impact of an auditory task on driving. The model is a good predictor of mental workload changes with auditory secondary tasks. However, predictions of the impact on performance from secondary auditory tasks were not demonstrated in the simulator study. Frequency of the distraction was more influential in the changes of performance and workload than the demand of the distraction, at least under the conditions tested in this study. While the workload numbers correlate with simulator numbers, using the model would require a better understanding of what the workload changes would mean in terms of performance measures.

15. SUBJECT TERMS

IMPRINT, human performance modeling, workload, driving, distractions

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Josephine Q. Wojciechowski
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED	SAR	75	**19b. TELEPHONE NUMBER** (*Include area code*) 410-278-8830

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Table of Contents

List of Tables

List of Figures

Acknowledgements

There are many people that have provided assistance to me so that I could complete a Masters in Industrial and Systems Engineering. I would like to take this opportunity to acknowledge most of them.

First, I need to acknowledge my committee members. Dr. Brian Kleiner was a member of my original committee and graciously stepped up to chair when I needed his help. He has been very supportive and provided exceptional guidance and assistance. Dr. Sue Hill from the Human Research and Engineering Directorate of the U.S. Army Research Laboratory was also on my original committee. She provided guidance on design of the experiment, use of the driving simulator, and lots of great input on the written documentation. Dr. Kari Babski-Reeves graciously joined my committee last spring when Dr. Beaton left Virginia Tech. She has been very helpful and provided some good suggestions. Lastly, I would like to thank Dr. Robert Beaton. Dr. Beaton was the original chair for my committee. He helped me to formulate the idea for this study and provided initial guidance in getting the proposal for this work finalized. Thanks to all the committee for keeping me working and focused.

Next, I need to thank the most important people in life, my family. Without their help, I would not have finished this. To my children, John, Joe, Matt, and Anna, thank you for understanding that mom had homework too. Thanks for always asking me when I was going to finish or asking, “Aren’t you done yet?” Watching you learn helped me to want to keep learning also. I am very proud of all of you. You are my joy. To my husband, Bob, I think I am going to finish this finally. Thanks for being so patient and encouraging. I certainly couldn’t have succeeded without your support. You are my inspiration. You are my soul mate. Thank you also for your service to this country. If it weren’t for you and all the members of the Armed Forces, these opportunities may not be available to us. I look forward to your homecoming. I think that I will be able to keep my promise to finish before you come home. To my mom, I feel so blessed that you have been able to be with me for the past year. You are so smart and inspirational. I hope that I make it to 91.

Finally, I would like to thank my co-workers and fellow graduate students. You all have kept me motivated and pressing on. It’s been a long road but I think I’m about to reach the rest stop. Thanks for keeping me company on this journey.

Chapter 1. Introduction

1.1 Background

Human performance modeling (HPM) is often used to assist in the design of systems. HPM is a particularly useful tool because system performance is always a function of the performance of the human operating the system. This is true irrespective of whether the system is a simple system such as a basketball and hoop or a complex system such as a jet aircraft. Task network models such as IMPRINT (Improved Performance Research Integration Tool) allow investigation of the human's impact on system performance (IMPRINT, 2004). IMPRINT is an example of a HPM that will be discussed in greater detail later. In many cases, HPM can be used for system analysis when it is not feasible to conduct a study. HPM can be used in situations where safety, cost, or practicability prohibits conducting a study of human performance on an actual system. HPM allows an analyst to examine system performance in terms of human performance. One of the most critical steps in applying HPM to design is verification and validation of the model. Validation is usually performed by comparison with data from field studies.

The terms, simulation and modeling, are often used interchangeably. Additionally, in this paper, a "simulator" is used as part of the experiment. There are several other terms that could be misinterpreted. To reduce confusion and ensure that the message is clear these terms are defined as follows for this work. Simulation is a tool that is used to create a representation of a system (in this instance, IMPRINT). Model is a specific representation of a particular system (in this instance, Driver Workload Model (DWM)). Simulator is an apparatus used to represent a system (Driver Workload Simulator (DWS)). Validation is defined as the extent to which the model is an accurate representation of the real world for the purpose for which it was designed. Workload, when used in this report, refers to mental demand.

A DWM was produced in IMPRINT, a HPM tool, to examine the difference between direct driving, teleoperation, and semi-autonomous driving (Wojciechowski, Kogler, and Lockett, 2001). The intent of this effort was to develop a model that included all the components and stages of information processing depicted in Wickens' information processing model (Wickens & Hollands, 1984). IMPRINT was chosen because of its ability to represent the human behaviors and the mental workload that characterize driving. The analysis of the model predicted that driver mental workload was high, that is, near or at the threshold of the operator's ability to maintain performance. This would indicate that any additional mental workload would reduce the performance of the driver.

Tasks associated with direct driving from the initial model were then combined with the additional tasks needed to operate a combat vehicle. The purpose of the new model was to investigate mental workload for different function allocations and crew size in a combat vehicle. The analysis from the combat model showed that the driver was consistently the crewmember with the highest mental workload and should not be required to perform any additional tasks above and beyond those related to driving (Mitchell, Samms, Henthorn, and Wojciechowski, 2003). It was deemed necessary to validate the DWM due to the importance of the driver.

Validation tests the accuracy of the model or simulation. "Validation is the process of determining the degree to which a model or simulation is an accurate representation of the real world from the perspective of the intended uses of the model or simulation." (DMSO, 2005) Validation gives one confidence that the results derived from the model output can be used to

answer the questions they were designed to resolve. In using a model to represent a system, it is understood that the model is just that, a representation. One does not try to recreate the actual system, as that would defeat the purpose of modeling. The goal in model building is to include enough detail to adequately represent the system for the purpose of the model. Validation is a means of determining if the representation adequately addresses the issue in question.

The DWM is being used to represent the driving tasks performed by soldiers in future combat vehicles; therefore, a valid model of driving is required. Driving tasks are critical to the operation of future Army systems. Initial modeling indicates that the driving tasks create perhaps the highest mental demand of all functions in the combat vehicles (Mitchell, et al., 2003). In order to insure that the driving tasks are a valid representation of the mental workload of driving, it is critical to collect data from experimental trials to validate the model conditions. While results from modeling efforts are helpful, validation of the results provides scientific data to insure the model is accurate for this purpose. The purpose of the DWM was to look at the mental workload of driving in conjunction with typical military tasks, such as listening to the communications radio. Therefore the DWM should be validated using driving behaviors with a secondary task to determine the accuracy of the outputs from the DWM with respect to mental workload and performance for the specific purpose for which it was and will continue to be used.

Validation of a model is a necessary and important process. Validation depends on the purpose of the model. It is completed "from the perspective of the intended use." (Department of the Army, 1997) In this case, validation implies that the representation of driving is appropriate for determining the mental workload associated with driving tasks in a military combat vehicle while performing additional tasks. When comparing the DWM constructs with other driving models, this model includes all the components of human information processing included in the other models. There do not appear to be incongruencies. This provides face validity to the structure of the DWM. Additionally, the driving and distraction studies showed that mental workload was at or near the threshold when driving, providing face validity to the outputs of the DWM. This is consistent with the results of this driving model in both IMPRINT studies (Mitchell, et al., 2003; Wojciechowski, et al., 2001). This implies that additional mental workload would result in the potential for performance errors. Based on these comparisons, it is believed that this representation of driving is valid for representing the mental workload associated with driving. However, further validation can be accomplished by comparison of the model output to empirical data. That is the purpose of this study. Driving tasks will be performed in a PC-based driving simulator rather than an on-road vehicle. This will avoid danger to participants in the event of performance errors. Drivers will be required to operate the driving simulator while completing a secondary task consisting of responding to auditory signals. The expectation is that this distraction will cause a decrease in performance and an increase in workload.

This work is valuable to the Army in design of any vehicle but primarily to the design of combat vehicles. Even if driving were automated, the visual and cognitive workload associated with monitoring or intervening in an autonomous mode would require that the operator be focused only on driving during some intervals. While technology advances are promising, current technology requires the full attentional demand of the driver. This driver model therefore is an important component in determining the functional allocation between crewmembers in military vehicles.

1.2 Research Questions

This investigation is designed to address specific research questions. Validation of a model must be performed in the context in which the model was or will be used. Therefore, the DWM must be validated for investigating mental workload during driving and performing additional tasks. The first step is to determine if the secondary task impacts the driving task and then determine if the HPM can predict the impact. To that end, the questions addressed by this investigation follow.

1. Does auditory distraction impact the performance and workload of driving as depicted in the DWM?

2. Can IMPRINT correctly predict the impact on performance and workload?

3. Is the DWM consistent with actual differences between less-demanding more frequent distractions and fewer higher-demand distractions?

1.3 Hypothesis

Listed below are the predicted answers to the previously stated research questions.

1. Auditory distractions while driving will reduce performance and increase mental workload.

2. IMPRINT can predict the change in performance and workload created when auditory distractions are imposed while driving.

3. More frequent, less-demanding auditory distractions will have more impact on performance and workload than less frequent, more-demanding auditory distractions in both the model and the real driving task.

1.4 Research Approach

The research approach utilized for this work is the model-test-model (MTM) approach as shown in Figure 1. MTM is listed as a technical approach to validation in DA Pam 5-11 (Department of the Army, 1999). In this approach, a predictive model is developed. The results of the model can lead to specific areas of concern and more focused testing. An experiment is then conducted that addresses the same system and conditions that are being investigated. The results of the model and the experiment are examined and compared to determine if the model is representative of the system. If necessary, results and variable relationships are used to refine model parameters and the model is again used to predict outcomes. This process can continue until the model parameters can predict system outcomes of interest. It is important in this process that any assumptions made in the development of the model are consistent with the assumptions made in the experiment. The model can only be used to predict the outcomes for model constructs and conditions that have been validated through actual testing.

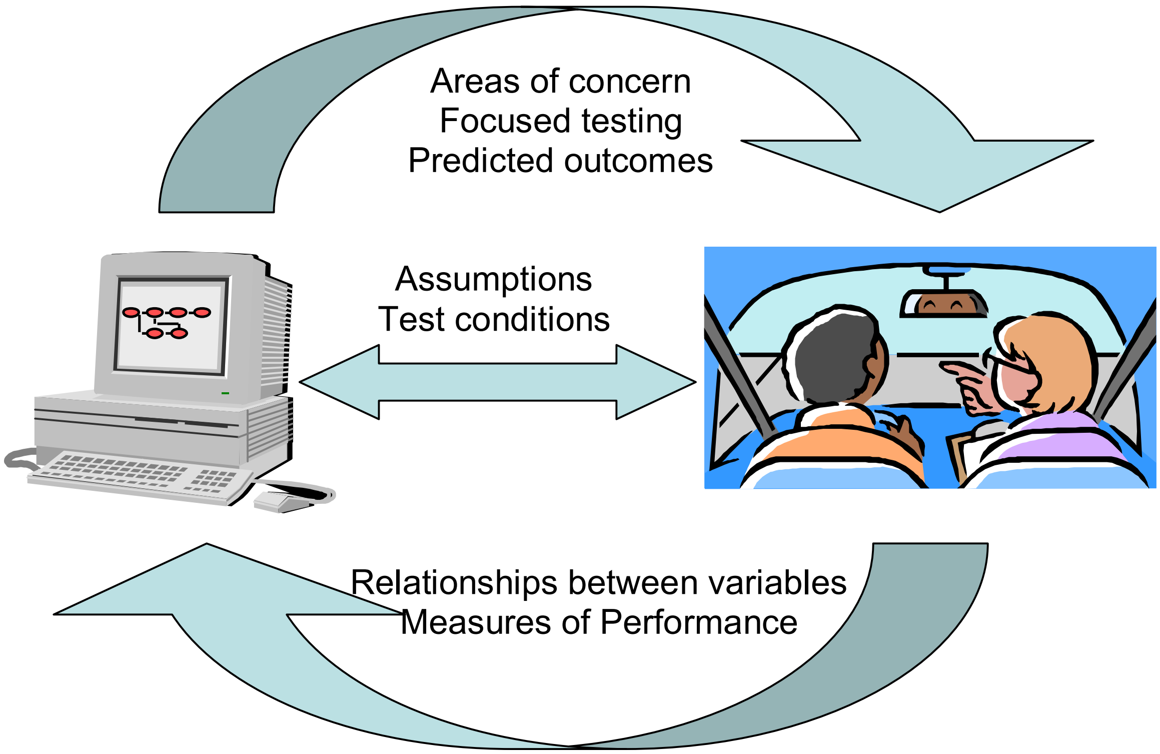

Figure 1. Model-Test-Model

Chapter 2 Literature Review

2.1 Use of Simulation

Two means to analyze a system are to conduct an experiment with the system or to model the system. The system can be modeled by simple mathematical models or by complicated computer software tools. The goal of modeling is to understand system performance. One uses a model to determine relationships between system variables in order to quantify system performance (Law and Kelton, 1991).

Human performance modeling is a means through which we can determine the human performance effect on system performance. System performance is always a function of the performance of the human operating the system. Therefore, it is important that the inherent variability of the human and its relationship to the system be understood. Variability of human performance is significant and therefore predicting how the human will perform is a challenge (Wojciechowski and Archer, 2002). Many simulation tools and strategies have been developed for this purpose. The U.S. Army Research Laboratory, Human Research and Engineering Directorate, developed a tool that can be used to measure the relationship between human performance and system performance.

Improved Performance Research Integration Tool (IMPRINT) is a task network simulation tool that can be used to answer many questions regarding human performance with respect to system performance. IMPRINT was developed in the 1970's from the common needs of the U.S. Armed Forces to understand manpower, personnel, and training requirements and constraints for proposed new weapons system. IMPRINT is constantly being updated and new capabilities added (Archer and Allender, 2001). It is a task network discrete event simulation tool with human performance algorithms and a graphical user interface built in to assist human factors analysts in system design and evaluation. IMPRINT can be used to set realistic system requirements, identify manpower and personnel constraints, evaluate operator and crew workload, test alternative system crew function allocations, determine maintenance man hours, assess performance under extreme conditions, and assess performance in terms of personnel characteristics and training (Archer and Adkins, 1999).

2.2 Validation

Validation is an integral part of modeling. A model should only be used to make decisions if it is a valid representation of the system. A model should be developed to address an issue or a set of issues (DMSO, 2005). The validation must also address this same issue(s). Validation provides the data to support the output or predictions of the model. Robinson (1997), however, reports that there is no such thing as absolute validity. The goal is not to show that the model is correct, but to show that it is not incorrect. This means that the model may not be an absolute representation of the system, but that it is correct for the context in which it is used. This will serve to increase the confidence in the model.

Validation should be performed throughout the life cycle of the model (Balci, 1997). There are many different techniques to validate a model. Where one is in the life cycle dictates which techniques are more useful. One of the first steps of any validation is validation of the task analysis used to create the model. This is often completed early in the life cycle of the model. One means of validating the task analysis is comparing the components of different driving models used to develop the DWM and determining that the task analysis was complete.

This can be completed by also by subject matter expert approval. Once the model is in a state to collect data, validation of results can be performed by comparison to data from other similar models or experiments. Formal validation includes mathematical proof of correctness for the purpose of the model (Balci, 1997). This is usually completed later in the life cycle of the model to show that the model can mathematically be used to address the purpose for which it was created. This study is an attempt to provide that proof for this representation of driving. It is an attempt to show that the predicted performance and workload resulting from this model can be used to represent driving in a combat vehicle where mental workload is the primary variable of interest. Previous experiments have been conducted aimed at validation of IMPRINT models (McMahon, Spencer, and Thorton, 1996; Mitchell, 1993). McMahon, et al. (1996) showed that an IMPRINT model can predict areas where performance of operators of nuclear, biological, and chemical reconnaissance vehicles would suffer. Mitchell (2000) showed that HPM can predict the performance of pilots in reconnaissance missions.

There is no prescribed process for validation. Validation is agreement between model results and the real system. The extent of the validation is dependent on the use of the model results. The results of this model indicate that distractions to driving or any secondary task in addition to driving (primarily visual and cognitive) would result in an increased potential for performance errors. These results are used to show how operators of military vehicles are impacted in terms of mental workload. Validation of this model by direct comparison of model outputs to actual driving data is a challenge, primarily because of the difficulty in measuring mental workload and the danger in exposing subjects to levels of mental workload that would lead to possibly dangerous performance errors. Therefore, validation was attempted by statistically comparing the results of the model to driving data collected on a simulator.

2.3 Mental Workload

Mental workload is the primary output measure in this model. Mental workload can be measured by many different means, such as primary task measures, secondary task measures, and subjective measures of workload (Sanders and McCormick, 1993). The difficulty is in comparing those empirical measures of mental workload with the output measures available in IMPRINT.

An important human information processing feature of IMPRINT is the capability to model mental workload demands. The mental workload demand theory, VACP (visual auditory cognitive psychomotor), implemented in IMPRINT is discussed in detail in McCracken and Aldrich (1984). This theory is based upon the notion that every task a human performs requires some attentional resources. All tasks have some level of mental demand, but it can vary widely depending on the task. Each task is comprised of demands in one or more of the VACP channels, such as visual or cognitive. IMPRINT is structured to help assign values representing the amount of demand by channel necessary to execute each individual task. IMPRINT uses a list of scale values and descriptors for each resource channel. These scales are taken directly from Bierbaum, Szabo, and Aldrich (1989). The scales are shown in Appendix A. Each scale ranges from 0.0 to 7.0 and has benchmarked textual descriptors corresponding to increasing demanding tasks in that channel. The descriptors correspond to increasing levels of human information processing activity within a given channel. The human information activity for each channel is considered to be the mental workload for that channel. When the model executes, each resource channel is summed for all tasks being prosecuted. This total value will represent the mental workload in that channel for the operator at that instance.

One difficulty in the comparison of the model data with experimental data resides in the mental workload measures. The VACP measures in IMPRINT are based on task demand and change during the course of the model run. The output from IMPRINT is a calculation of workload-over-time by mental resource. The analyst has the ability to set workload thresholds by channel or by some mathematical calculation of the individual channel scores. Typically, workload thresholds are set at greater than seven (7) for any individual channel or greater than 40 for a simple sum of the four channels (justification follows). As the operator performs different tasks, workload is measured and the total time that drivers are over a workload threshold is calculated as percent of time in overload.

Workload threshold of 40 was chosen based on previous studies and simple mathematics. A workload threshold of 40 is traditionally used as a threshold in VACP studies (Mitchell, 2003; Mitchell, et al., 2003; Wojciechowski, et al., 2001). Each of the workload scales reaches a maximum at 7.0. There are four scales that are applied to each task. If one assumes that each scale is at its maximum for a task, the overall workload or sum of the channels is 28. However, a single score of 28 does not mean that all scales are at a maximum. It is possible, for example, that in the concurrent tasks when the workload is added, cognitive and visual channels are above 7.0 and psychomotor and the auditory channels are lower. Therefore, a result greater than 28 would indicate that at least one channel was at the maximum in more than one task. A workload score of 40 in IMPRINT would mean that several channels are overloaded or at least one channel was severely overloaded in more than one task. In a condition where several channels are overloaded or one severely overloaded, the probability of performance errors is likely to increase. In the DWM, the number of tasks that could be executing simultaneously ranges from two to eight.

The National Aeronautics and Space Administration-Task Load Index (NASA-TLX) is a subjective assessment of workload (Hart and Staveland, 1988). It has been used in many different domains for measuring workload on operators performing their jobs. The validity of this tool for measuring workload and predicting performance has been demonstrated in several cases (Hill, Iavecchia, Byers, Bittner, Zaklad, and Christ, 1992; Rubio, Díaz, and Martín, 2004). NASA-TLX has recently been used to measure workload while driving with distractions (Slick, Cady, and Tran, 2005). The NASA-TLX tool rates workload on six subscales; mental demand, temporal demand, physical demand, effort, frustration and performance. Each subscale is a unidimensional scale rating from 0 (very low) to 100 (very high). The resultant scores represent a subjective measure of the perceived workload of the operator. NASA-TLX subscale scores can be evaluated separately or averaged. Also, a paired comparison of the subscales can be used to weight the subscales before averaging. However, Byers, Bittner, and Hill (1989) suggest that omitting the paired comparison procedure in NASA-TLX will not compromise the outcome of the measure.

2.4 Driver Models

The DWM is described and other driver models are compared for validation of the concept model.

2.4.1 DWM Model Description

The DWM is a simple IMPRINT model in that it does not involve a large number of tasks. It includes the primary mission, to drive from point A to point B. This mission consists of three functions: “Move”, “See”, and “Maintain Situation Awareness (SA)”. These functions all

run concurrently in the model. They represent the informational processing involved in driving. Each of the functions is described separately and shown in Appendix B.

The "Move" function includes tasks that represent steering and controlling the speed of the vehicle (acceleration, deceleration, and coasting). These tasks are set to occur in a cyclical fashion, meaning that once the vehicle has initially accelerated, a probabilistic decision is made whether the driver will accelerate, decelerate or coast. Once this task is complete, the probabilistic decision is executed again with the choice to speed up, slow down, or maintain speed. Initial speed is an input and can be changed. The increase or decrease in the acceleration and deceleration tasks respectively can be set to the desired level. A minimum and maximum speed can also be set. The driver also cycles through the steer and do-not-steer tasks. As each task is performed by the operator, the mental demand associated with performing that task is calculated from VACP ratings.

The "See" function represents the tasks associated with the perceptual and decision making portions of information processing. Included in this function are the tasks of scanning the sector, detecting landmarks, recognizing the path, calculating the distance to objective, and comparing to the guidance. Initially, the driver scans the sector. Then, the model will probabilistically determine if the driver sees a landmark. The driver will then continue his tasks whether a landmark is detected or not. The following task is recognizing the path. Then, calculating the distance to the objective and comparing to guidance received are performed simultaneously. These three tasks, recognizing the path, calculating distance to the objective, and comparing to guidance, represent cognitive processes while scanning the sector represents the perceptual process. That is why scanning the sector is executed before the three cognitive tasks. These tasks repeat to form a feedback loop similar to that described in Wickens' human information processing model (Wickens and Hollands, 2000). While the perceptual and cognitive tasks are separate in the model, IMPRINT can account for visual and cognitive workload in both types of tasks. In the perceptual tasks, the visual workload is high and the cognitive workload is lower. In the cognitive tasks, cognitive workload is higher and the visual workload is reduced as compared to the perceptual tasks. This allows for continuous but varying demand in both resource channels.

Another function that repeats is the "Maintain Situation Awareness" function. This function was originally built into the model as a direct result of the cognitive task analysis performed to develop the task flows for the model. This function consists of cognitive processes that include assessing the orientation of the vehicle, assessing the motion of the vehicle, assessing the traction of the vehicle, and awareness of vehicle function. For an experienced driver, these tasks are learned and are performed almost automatically based on cues from the environment (Schlegel, 1993).

The DWM is designed to address the mental workload of the tasks that occur during driving. These tasks are presented in a pseudo-random order, that is, each task begins based on logical sequences build in the model. However, because of the stochastic design of IMPRINT, one cannot predict which task will execute concurrently with other tasks.

The parameters of the model were designed so that the model would probabilistically replicate the parameters of the course represented in the driving simulator. The course was divided into sections. Each section represented a basically straight path or a curve. A goal speed for each section was determined based on the length of the section and radius of the curve. The model would compare the current speed with the goal speed for that section of the course and probabilistically accelerate, decelerate, or coast depending on the difference.

Functions, tasks, and goals for the model were developed by employing hierarchical task analysis (Kirwan and Ainsworth, 1992) and then augmented by using cognitive task analysis to capture the non-physical aspects of driving and controlling vehicles (Cooke, 1994). These methods are a means of developing a task analysis for a system in order to insure that all aspects of the operator's interaction with the system are considered. Hierarchical task analysis led to separating driving into three primary functions, psychomotor tasks, visual and cognitive tasks, and other sensory tasks. The psychomotor tasks included steering and controlling speed (i.e. accelerating, decelerating, braking or maintaining constant speed). The visual tasks included searching and deciding on the path. The other sensory tasks included situation awareness tasks such as listening and feeling changes in the vehicle. Cognitive task analysis allowed the inclusion of specific tasks that would incur mental workload. Task duration times not covered by modeling assumptions were developed based on data and algorithms found in the literature (Wierwille, 1993; Archer and Adkins, 1999).

2.4.2 Other Driver Models

There are many existing models of driving. It is important to note that each model is built to answer a specific question or set of questions. For this reason, different models of driving are created for each type of investigation. Any human performance model should however begin with a task analysis of driving.

Levison (1993) described a "Driver Performance Model" that was developed in 1993 and has since been used as a basis for other driving models. The parts represented in Levison's model include perception, cognition, control actions, and decision-making. This model is actually two models combined, a driver/vehicle model and a procedural model. The driver/vehicle model is a continuous feedback model between the driver's actions and the vehicle reactions. The procedural model looks at the driving tasks and determines task selection along with modeling the in-vehicle auxiliary tasks. The procedural model represents the regulation of attention. These components are all represented in the DWM.

Biral and Da Lio (2001) suggested that good driver models are required to predict vehicle performance. Their investigation revealed three main types of driver models. First, some models are based on conventional continuous control such as Proportional Integral Derivative (PID) and Generalized Predictive Control (GPC). The second type of driver models that exist are fuzzy logic or neural network based controllers. Fuzzy logic controllers are popular for representing human behavior and neural nets for their capability to learn. The final class of driver model that can be found are called hybrid and hierarchical models. These make use of the other two previously described types. Of the driver models Biral and Da Lio identified, they determined that for models to represent realistic driving behaviors the models must functionally consider the following components, perception, cognition, decision, and motor process of the human.

Salvucci, Boer and Lui (2001) use a cognitive architecture to model driver behavior. They characterized their model in terms of three primary components: control, monitoring, and decision-making. The control component accounts for perception of control variables and motor control. The monitoring component accounts for monitoring the environment. The decision-making component is the cognitive process of determining if a lane change is necessary or safe.

Brown, Lee and McGehee (2000) described a driver model of rear-end collision warnings. The results are a time history of the driver's response in avoiding a rear-end collision.

Their model contains three major components. The first is a representation of the attention to the roadway based on the uncertainty of the driver. The second component describes the decision process for braking or travel. The third component describes the driver's response. Again, these are the perceptual, cognitive (to include decision making), and motor processes.

Additionally, there has been some discussion about the adequacy of representing a continuous process (driving) with a discrete event simulation (ARL TAB, 2002). The most continuous portion of driving would be the visual and cognitive processes. Even these can be described as discrete tasks. The continuous task is divided into discrete chunks with no time interruption in the simulation clock. This represents a "continuous" process. Harrell and Tumay (1997) state that it is possible to model continuous phenomena using discrete-event logic, particularly when a high degree of precision is not required. They also state that discrete change systems can be modeled using continuous simulations when the state changes at small intervals. In this case, both mental workload and performance measures are not required at a high level of precision, it is believed that representing driving as a discrete event process is appropriate.

2.4.3 Comparison of DWM to Other Models

All the human information processes represented in each of these other models are also represented in our discrete-event simulation model. The representations may be different but this is expected because the purpose for each of the driver models is different. Most driver models are built in a closed loop system with the vehicle so that the actions taken by the driver model will impact the vehicle performance and that will in turn impact the next action of the driver. The DWM was actually built to determine the attentional demands that are controlled in Levison's (1993) procedural model. The feedback loops with the vehicle are represented in the DWM by probabilistic decisions. The DWM is a stochastic model used to look at the different combinations of driving tasks that may happen concurrently. This provides the ability to identify how the driver's mental demand varies and identify areas for potential performance degradation.

The output from the DWM suggests the potential for performance errors while driving is great. There are many times in a model run when the driver's mental workload is near or above what might be considered a mental workload threshold, as previously discussed in section 2.3. This would indicate that any distraction to driving would increase the probability of performance errors.

2.5 Driving and Distraction

Driver distraction and subsequent performance errors impact the safety of the vehicle. Performance errors can include a range of errors from simple ones such as lane variation, speed reduction, and missed traffic signs to more serious errors such as accidents. Therefore, insurance companies, automobile manufacturers, government agencies and other policy makers are all interested in the topic of driver distraction. As a result, many studies have been conducted to quantify and qualify the performance errors that may be caused by different driver distraction. Studies have been conducted in instrumented vehicles and in simulators. Almost all investigations show that any distraction to driving provides the potential for performance errors. Discussion of several of these studies follows.

Cell phone use is one of the most common distractions to driving that has been studied recently. In 1997, the epidemiological study of Redelmeier and Tibshirani reported in the *New England Journal of Medicine* concluded that cell phone use quadrupled the risk of collision during the period of the call. Strayer, Drews and Johnson (2002) performed a series of

experiments in simulators that showed that talking on a hands-free cell phone while driving caused what they labeled "inattentive blindness". The experiments ranged from observing driving performance errors to determining that drivers do not recall billboards that were fixated on while driving and talking on the cell phone. Direct Line Motor Insurance (2000) has shown that reaction times for drivers were on average 30% slower when the driver was engaged in a cell phone conversation and driving than when the driver was legally over the limit for alcohol consumption and driving. Furthermore, the reaction times for drivers talking on a mobile phone were 50% slower than when they were driving without talking on the phone.

Driver distraction is a large research area. Tijerina (2000) reported that predicting costs and benefits of driver distraction associated with in-vehicle technology is very complex and difficult. However, driver behaviors and operational problems with the technology can be evaluated. There is evidence that crash data and driver distraction are related. There are, however, so many variables that it is difficult to predict what level of distraction would cause an accident. Tijerina uses an analogy about smoking and lung cancer. You will not necessarily get cancer from smoking but the risk is much greater. Similarly, you may not have a performance error if you are distracted while driving, but the risk of error is much higher.

Chapter 3 Methodology

Validation of the model was attempted by comparison of model data with performance and workload data from a human-in-the-loop driving study using a desktop simulator. Two experiments were conducted. The first experiment consisted of IMPRINT model runs. The second experiment was comprised of a human-in-the-loop driving simulator study. Both the modeling and simulator study resulted in measures of driving performance and workload while attending to a secondary task. The model runs were generated separately and prior to the simulator data collection. Modeling results and simulator experiment results were then compared.

3.1 Experiment Design

Each experiment (model runs and simulator study) was run as a two factor within-subjects experiment. One factor was complexity of secondary task. The other factor was the frequency of the secondary task. While the model did not have any subjects, the base model was run in each condition with only the complexity and frequency of the secondary task changing. In the simulator study, each subject performed all conditions. The model results are called Experiment 1 and the human-in-the-loop simulator study is called Experiment 2.

3.1.1 Independent Variables

Two independent variables were used in this study. The first independent variable was secondary task complexity. Complexity of the secondary task was represented by required participant responses to auditory signals. The complexity was varied by increasing the number of signals to which the participant was required to respond. Participants were subjected to an auditory signal. This signal was a word. The participant was instructed to respond with a specified response for that signal. For example, if the signal was the word “one,” the participant should respond with “red.” The complexity levels are listed below.

1. Driving with no distraction.
2. Driving with one response to the trigger signal.
3. Driving with two separate responses to two different trigger signals.
4. Driving with four separate responses to four different trigger signals.

The signals used were four words; “one,” “two,” “three,” and “four.” The response to “one” was “red.” The response to “two” was “white,” the response to “three” was “blue,” and the response to “four” was “green.” In factor level 2, only one signal was used (i.e. “one”). This level represented an automatic response with little mental processing. In level 3, the signals “one” and “two” were used. In level 4, all four signals were used. These two factor levels (3 and 4) represented a requirement for higher level mental processing. In levels 3 and 4, the various signals were randomly presented. The response for each signal was consistent in all trials.

The second independent variable was the frequency of the signals. Two levels of frequency of signal were presented. The levels were two signals per minute and six signals per minute. At the two-signal per minute level, the signals were presented randomly within each 30 seconds. At the six-signal per minute level, they were presented randomly within each 10

seconds. A diagram of the conditions presented is shown in Table 1. The secondary task conditions shown in Table 1 were the independent variables.

Table 1. Experiment design

<table>
<tr><td colspan="2" rowspan="2">C = condition number
n = number of subjects (model runs)</td><td colspan="4">Number of Signals</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>4</td></tr>
<tr><td rowspan="2">Frequency of Signals (# per minute)</td><td>2</td><td>C1, n=14(50)</td><td>C2, n=14(50)</td><td>C3, n=14(50)</td><td>C4, n=14(50)</td></tr>
<tr><td>6</td><td>Same as C1</td><td>C5, n=14(50)</td><td>C6, n=14(50)</td><td>C7, n=14(50)</td></tr>
</table>

3.1.2 Dependent Variables

The dependent measures for Experiment 1 were mental workload and performance measured by mission completion time and average speed. Workload was measured as percent time in overload in experiment one and subjective workload in the second experiment. Additionally, the average response time, or the time between the auditory signal and the operators' response was calculated. Dependent variables for both experiments are shown below and defined in Table 2.

1. time to complete the course
2. average speed
3. workload under each condition
4. response time

Additional data collected included the number of correct response to signals. No response or a response later than 5 sec was considered an incorrect response. IMPRINT has micromodels of human performance times built into the tool based on times that are published in literature (Archer and Adkins, 1999). This level of 5 seconds was chosen because the IMPRINT micromodels show that choice reaction time for four choices is calculated at 0.35 seconds (Card, Moran, and Newell, 1983). Speech rate for one word is 0.34 seconds (McCormick, 1970). Therefore, allowing 5 sec for a response was enough time to consider a correct response. This data was not analyzed as the model did not predict any incorrect responses and there were only 20 incorrect responses for all subjects under all conditions for the human-in-the-loop study.

3.2 Experiment 1 - IMPRINT model runs

The DWM was used to represent driving tasks. The general DWM can be adjusted to examine specific driving courses or specific additional tasks. A specific auditory secondary task was added to the model for this study. The secondary task was represented in the model by several tasks. The operator would hear the signal in one task. The task that followed was the response task. Workload for the response task and time to respond varied with the number of responses from which to choose. These two tasks were the only ones that represent actions of

the operator. The coding in the DWM was manipulated to characterize different complexities and frequencies of the secondary task as displayed in Table 1.

Table 2. Dependent variable definitions

Dependent Measure	Model Definition (Experiment 1)	Simulator Study Definition (Experiment 2)
Time to Complete the Course	The length of time from start to finish, the mission completion time.	The length of time from start to finish, the time it took to drive the complete course.
Average Speed on the course	The average speed driven on the course, average speed is calculated by distance traveled divided by the mission completion time.	The average speed driven on the course, average speed is an output from the simulator and represents sum of all the speeds for each screen update divided by the number of screen updates.
Response Time	Response time is the delay between the auditory signal and the response, based on the micromodels built into IMPRINT. The mean response time is determined by the micromodels and a distribution of times around that mean is used. The output is the stochastic time chosen in each instance from that distribution of times.	Response time is the delay between the auditory signal and the response; this was measured using a MP3 recorder. The recordings were analyzed using the Sound Forge software to determine time between the signal and response.
Workload	% Time in Overload, percentage of the total time the operator's overall workload was greater than 40.	NASA-TLX subjective workload score, average of all six individual scales.

3.2.1 Participants

There were no human participants required for this investigation.

3.2.2 Apparatus

A personal computer and IMPRINT software were the only apparatus required for this experiment. The DWM was created in IMPRINT as shown in Appendix B. The simulation package IMPRINT is available from the U.S. Army Research Laboratory (IMPRINT, 2004).

3.2.3 Procedure

The DWM was modified to add the secondary tasks as discussed above. The model was executed the same number (50) of times in each condition. Data from each condition

(complexity of signals and frequency of signals) was analyzed for performance measures and mental workload according the experiment design.

The number of model runs was based on the variability of the model. IMPRINT is a stochastic modeling tool. There are two sources of variability in the model. One source is the variability that results from the process modeled. The other is based on the random number seed that is chosen to initiate the model. In order to eliminate the variability from the random number seed, the model must be run multiple times with different random number seeds. To determine the correct number of runs, one condition was run 10, 20, 30, and 50 times with three different random number seeds. The overall mission times were examined to determine how many model runs are needed for the variability to stabilize. At this point, it was determined that the variability in the output was due to the conditions of the model and not to the choice of random number seed. Fifty was a sufficient number of model runs for the variability to level off.

The output measures for each of the 50 runs were used in the analysis. Each model run produced a mission completion time, an average speed and percent time in overload. The response time was the mean of the response times for that run. This produced fifty data points for each dependent measure per condition.

3.2.4 Data Analysis

IMPRINT records the output from the model in results files that document when every task starts and stops, the current workload for each channel, and the task information. The analyst can create "snapshot" files that record specific data of interest. All data from the model was collected in these output files from the simulation software. Excel® (MicroSoft, version 2003, 2003) spreadsheet macros were built to format the data in the IMPRINT output files into the format needed to process these files in SPSS® (SPSS, version 12.0, 2003). These data were evaluated using ANOVA as described in the results section.

3.3 Experiment 2 - Simulator Study

Once the model runs were complete, the same tasks were conducted in a driving simulator to measure operator performance and subjective workload.

3.3.1 Participants

There were fourteen participants who completed this study. The participants were 13 volunteer civilian employees with one volunteer military subject. Volunteers were recruited by personal invitation or email. All participants were 18 years old or older.

Prior to the research, each participant was briefed on the research and asked to provide his or her informed consent to participate. This included an explanation of the purpose and procedures as described in the Volunteer Agreement Affidavit.

Each participant was asked if he or she had any medical injury or condition that would preclude him or her from participating. Any potential participant with medical history or concerns about the study was not allowed to participate.

Each participant was trained on the driving task. The participant completed the number of trials needed for the participant to feel comfortable with the course. Participants were then trained on the secondary task. Participants were also trained on the NASA-TLX questionnaire.

Two subjects did withdraw from the study due to motion sickness symptoms. Both participants left during the training phase, prior to any experimental trials. No data from these two participants were used in the analysis or results.

All data collected in the study was coded by participant number (i.e. no names were used) and kept confidential. At the end of the study, each participant was debriefed on their results and any other questions they had were answered.

3.3.2 Apparatus

3.3.2.1 Driving Task

For the driving task, participants used a desktop simulator. The simulator was constructed from a Dell OptiPlex GX400 personal computer. A steering wheel, accelerator, and brake pedals (NASCAR Pro Digital 2 by Thrustmaster) were used by the participants to determine direction and speed. The software for the Driver Workload Simulator (DWS) was developed for the U.S. Army Research Laboratory by Shankle (2002). This simulator was originally built to increase the capability to test system readiness and technical maturity in a synthetic environment. This software is a representation of one of the U.S. Army's Aberdeen Proving Ground driving test courses. It presented a brown path on green grass for dirt roads and gray path for paved roads. The course was delineated by road signs and barriers at intersections and cross paths. Sharp curves were preceded by road signs indicating the direction of the bend. Figure 3 show an example of a screen shot from the simulator. In this screen shot, a truck is placed in the scene, off-road. A gun tube is present as part of the simulated vehicle the participant is driving.

3.3.2.2 Distraction Tasks

The secondary task was presented by the speakers on the Dell personal computer. A .wav file was created for each condition that presented the defined number of signals at the defined frequency for that condition. The .wav file was initiated when the participant began driving. At the same time, the sound in the lab was recorded to collect both the signal and the response using a Rip Flash Trio Digital Voice Recorder (EVR-100). This recorded sound file was used to determine the time between the signal and the response. Sound Forge® (Sony, version 4.5, 2003) was used to evaluate the recorded sound files for response times.

3.3.2.3 Questionnaires

Questionnaires were used to record subjective workload and demographic data. NASA-TLX data sheets collected workload in six subscales. The demographic questionnaire was used primarily as a screening tool for motion sickness problems. Questionnaires are provided in Appendix C.

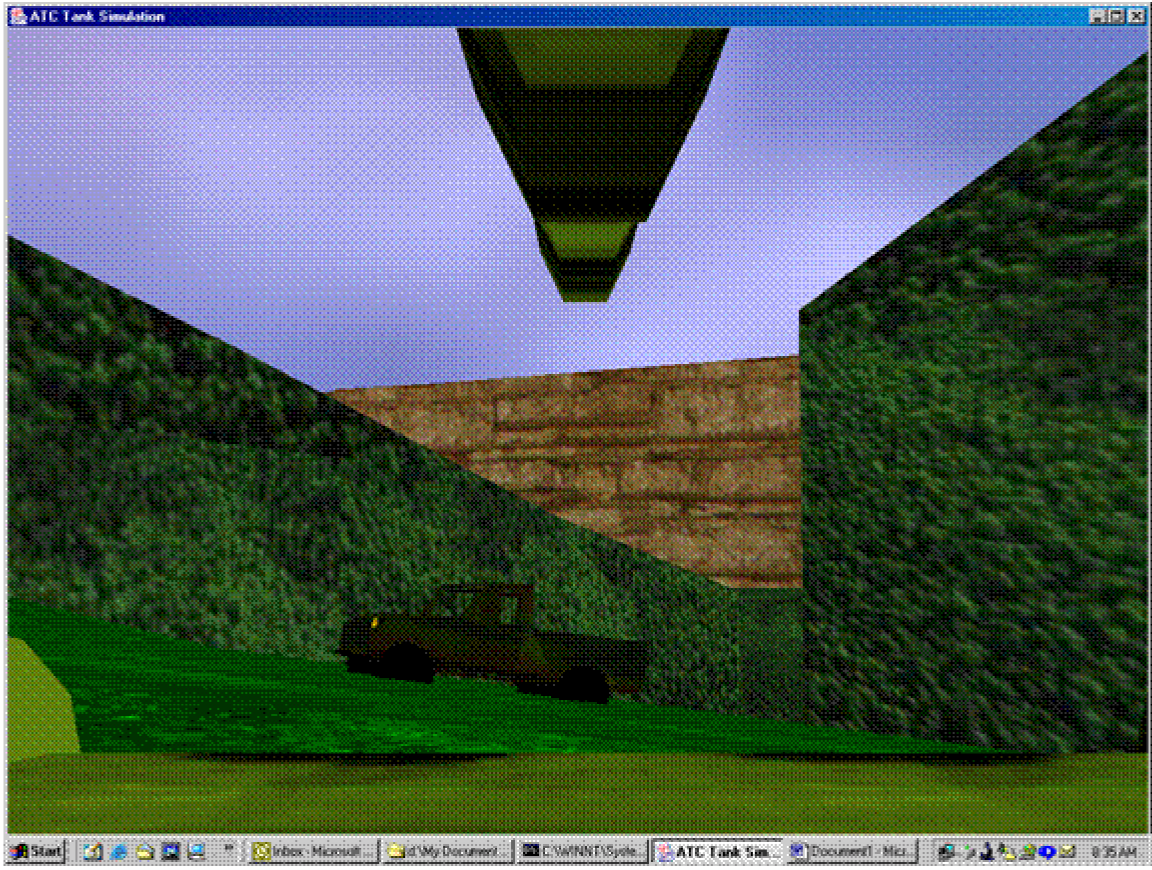

Figure 2. Driver's view in the DWS simulator

3.3.3 Procedure

In order to account for order effects, the levels of these factors were presented using a Williams Design shown in Table 3 which counterbalanced the presentation of conditions. One participant was evaluated at a time. Each participant performed the driving task under all secondary task conditions and frequencies. All conditions were tested at one session.

The participant reported to the test room which was an office. The office contained a desk, a table, and a credenza with two chairs, one for the participant and one for the experimenter. The participant used the desk for filling out all paperwork including the volunteer agreement and the questionnaires. The DWS was set up on the table opposite the desk.

The experimenter described the test during the orientation. All tasks were explained in detail. The participant was given the opportunity to ask and have answered any questions pertaining to the test and their participation. He/she was then asked to read and sign the Volunteer Agreement Affidavit. The demographic questionnaire was filled out next.

Table 3. Williams design for experiment conditions

	Order						
Subject	1	2	3	4	5	6	7
1	C1	C7	C2	C6	C3	C5	C4
2	C2	C1	C3	C7	C4	C6	C5
3	C3	C2	C4	C1	C5	C7	C6
4	C4	C3	C5	C2	C6	C1	C7
5	C5	C4	C6	C3	C7	C2	C1
6	C6	C5	C7	C4	C1	C3	C2
7	C7	C6	C1	C5	C2	C4	C3
8	C4	C5	C3	C6	C2	C7	C1
9	C5	C6	C4	C7	C3	C1	C2
10	C6	C7	C5	C1	C4	C2	C3
11	C7	C1	C6	C2	C5	C3	C4
12	C1	C2	C7	C3	C6	C4	C5
13	C2	C3	C1	C4	C7	C5	C6
14	C3	C4	C2	C5	C1	C6	C7

The next step was completion of the training trials. The participant was introduced to the equipment and the procedure for operating the DWS. The operating instructions were scripted in order to insure that all participants were given the same information. The participant was instructed to proceed through the designated course at a fast but comfortable speed while attempting to keep the vehicle on the test course. The participant was told that driving accuracy takes precedence over time. The training trials continued for four trials or until the participant felt comfortable with the course. The first two training trials were conducted at 15 mph maximum speed. The next two trials were conducted at 40 mph maximum speed. This was the maximum speed during the experimental trials; the maximum speed can be set by the experimenter. Only one participant wanted to complete an additional training trial. The participants were then trained on the secondary task. Participants were given a “cheat sheet” with each signal-response combination on it for the beginning of this trial. The participant responded to signal words for 6 minutes at 1 word every 5 seconds or 12 words a minute. The words were presented in order twice. Thirty seconds later, the “cheat sheet” was removed and the participant was required to respond from memory for the remainder of the six (6) minutes. Participants generally answered with the correct response. They averaged 1 incorrect response in the 6 minute block. At that time, the participant was asked to complete the workload survey to measure workload of the secondary task alone. The participant was then asked to complete an additional driving trial. In this training trial, signals were introduced and the participant was asked to respond appropriately when a signal was detected. At the conclusion of these training trials the participant was allowed to rest for 5 minutes. On average, the training lasted approximately one hour.

The experimental trial began next. Each participant was asked to drive on a course for a specified distance. Each participant drove the simulator under each of the secondary task conditions. The order of presentation of these conditions followed the Williams Design to

counterbalance for order effects. At the conclusion of each driving trial, the workload assessment for that trial was administered. The participant was allowed a 10-minute rest between each experimental condition. At the conclusion of all the experimental trials the participants were released for the day. The total time was approximately four hours.

3.3.4 Data Analysis

The simulator automatically collected data from each trial for time on course and average speed. Signal response audio data was recorded on a digital MP3 recorder. The MP3 files were converted to .wav files and analyzed with Sound Forge®. Response time was measured as the time from the end of the signal peak to the beginning of the response peak. NASA-TLX data was collected on a paper survey. NASA-TLX subscale data was averaged for an overall score. The scales on NASA-TLX were equally weighted based on Byers, et al. (1989). In this instance, the task conditions were so similar that it was not likely the paired comparisons would differ from condition to condition.

3.4 Comparison Analysis

The results from both the model and the simulator study were analyzed. Since the model is a stochastic model not a deterministic model, it does not predict exact performance and workload. The results are a predicted average of the possible results given this system and the variability that is represented in the model. Therefore, the simulator data and the model data were compared to show that they correlate. Correlations were completed on all the dependent measures from both experiments.

Chapter 4 Results

As stated in Introduction, the goals of this research were to answer the following questions:

1. Does auditory distraction impact the performance and workload of driving?
2. Can IMPRINT correctly predict the impact on performance and workload?
3. Is there a difference between less-demanding more frequent distractions and fewer higher-demand distractions?

Analysis of the data from these two studies was completed using the SPSS© statistical software package. For each experiment separately, dependent measures were analyzed by an ANOVA. All post hoc analyses were completed using the Least Significant Difference (LSD) method. The experiment design was a 4 x 2 factorial. The two independent measures were number of signals and frequency of signals. At zero level of number of signals, frequency did not matter (zero signals at two times a minute is the same as zero signals at six times a minute). That gives seven conditions or a 3 x 2 matrix with a control condition. The ANOVAs were conducted by first doing ANOVA on the seven conditions to determine the significance of the conditions from one another and to calculate the degrees of freedom and mean sum of squares for the error term. A second ANOVA was completed on the six non-zero conditions (3 x 2 factorial) to determine the significance of the individual factors and their interactions. The final table was calculated using the error term from the first ANOVA. In the second experiment, subjects and order were also considered in the ANOVA.

4.1 Experiment 1 – Model Results

The results of the first experiment follow. The dependent measures that were collected included time on course, average speed, and response time for the secondary task. Also, the percent time in overload was recorded for each trial.

The descriptive statistics for time to complete the course are shown in Tables 4 and Figure 4. ANOVA results are shown in Table 5. For Experiment 1, ANOVA results indicated a significant main effect for both condition and signal frequency ($F_{6, 343}$=2.964, $p \leq 0.01$ and $F_{1, 343}$=11.140, $p \leq 0.001$, respectively) for time on course. Post-Hoc tests indicated significant differences between Conditions 1 and 2 (p=0.002), Conditions 1 and 3 (p=0.007), Conditions 1 and 4 (p=0.005), Conditions 2 and 5 (p=0.042), Conditions 2 and 6 (p=0.040), Conditions 2 and 7 (p=0.020), and Conditions 4 and 7 (p=0.038). These are listed in Table 6.

Table 4. Time to complete the course for model conditions (minutes)

Condition	Mean	Std. Dev.
No signals	6.17	0.35
1 signal – 2 per minute	6.39	0.31
2 signals – 2 per minute	6.36	0.31
4 signals – 2 per minute	6.37	0.36
1 signal – 6 per minute	6.25	0.34
2 signals – 6 per minute	6.25	0.31
4 signals – 6 per minute	6.23	0.43

Table 5. ANOVA for time to complete the course for model conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Condition	6	0.357	2.964*
Signal Frequency	1	1.348	11.140**
Number of Signals	2	0.010	0.083
Signal frequency * Number of signals	2	0.007	0.057
Error	343	0.121	

* $p \leq 0.01$
** $p \leq 0.001$

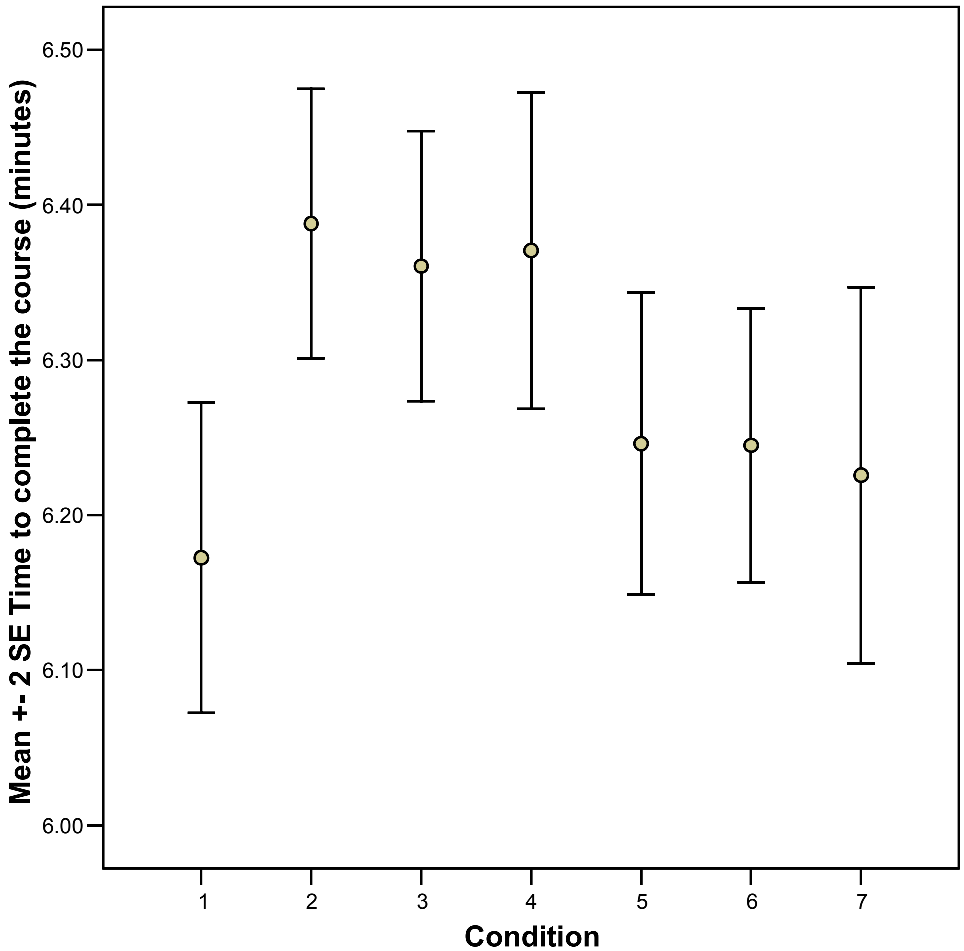

Figure 3. Time to complete the course for model conditions (minutes)

Table 6. Significant differences between conditions for time to complete the course in the model study

Condition	**Significantly Different Conditions**
No signals	2,3,4
1 signal – 2 per minute	1,5,6,7
2 signals – 2 per minute	1
4 signals – 2 per minute	1,7
1 signal – 6 per minute	2
2 signals – 6 per minute	2
4 signals – 6 per minute	2,4

Descriptive statistics for average speed on the course are presented in Table 7 and Figure 5. ANOVA results are presented in Table 8. ANOVA results indicate a significant main effect for both condition and signal frequency for average speed on course ($F_{6, 343}$=2.936, p≤0.01 and $F_{1, 343}$=11.240, p≤0.001, respectively). Post hoc tests indicated significant difference between Conditions 1 and 2 (p=0.003), Conditions 1 and 3 (p=0.008), Conditions 1 and 4 (p=0.006), Conditions 2 and 6 (p=0.044), Conditions 2 and 7 (p=0.014), Conditions 3 and 7 (p=0.035), and Conditions 4 and 7 (p=0.029). The significant differences between conditions are listed in Table 9.

Table 7. Average speed on the course for model conditions (miles per hour) (Max allowed speed was 40 mph)

Condition	Mean	Std. Dev.
No signals	30.39	1.83
1 signal – 2 per minute	29.36	1.45
2 signals – 2 per minute	29.48	1.46
4 signals – 2 per minute	29.45	1.67
1 signal – 6 per minute	30.02	1.71
2 signals – 6 per minute	30.05	1.54
4 signals – 6 per minute	30.20	2.17

Table 8. ANOVA for average speed on the course for model conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Condition	6	8.539	2.936*
Signal Frequency	1	32.698	11.240**
Number of Signals	2	0.470	0.162
Signal frequency * Number of signals	2	0.209	0.072
Error	343	2.909	

* p≤0.01
** p≤0.001

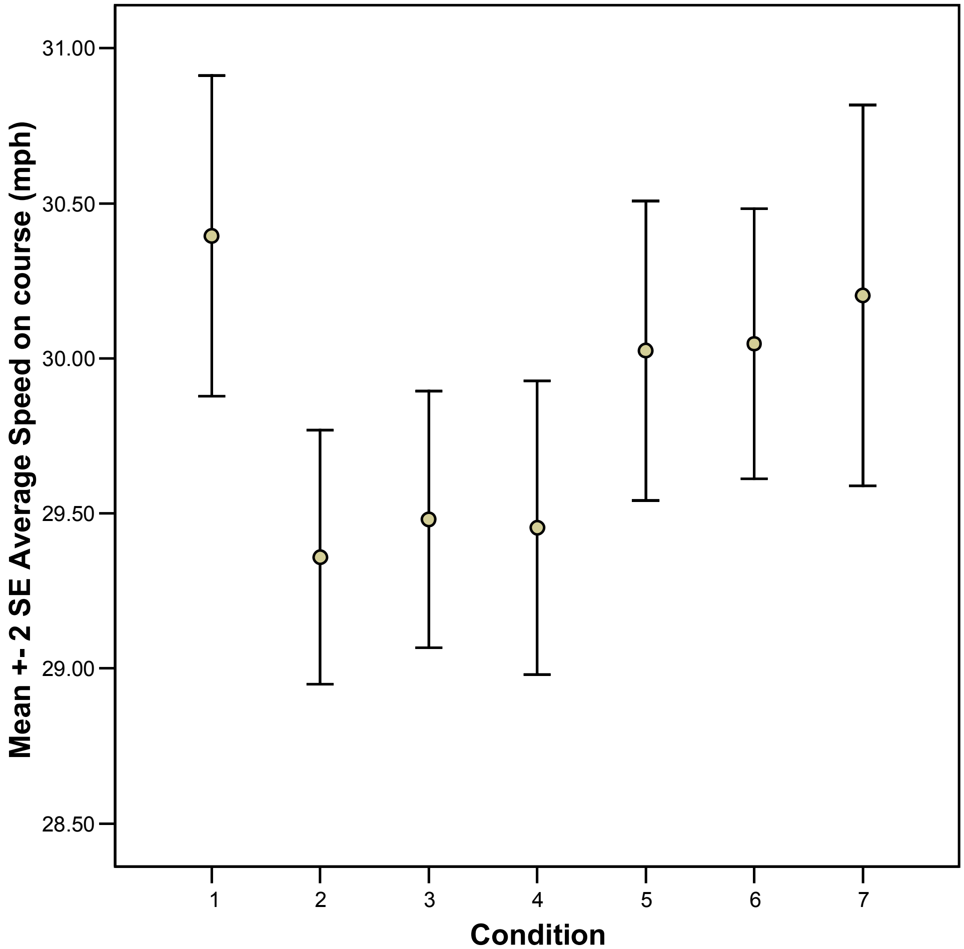

Figure 4. Average speed on the course for model conditions (miles per hour)

Table 9. Significant differences between conditions for average speed for the model study

Condition	**Significantly Different Conditions**
No signals	2,3,4
1 signal – 2 per minute	1,6,7
2 signals – 2 per minute	1,7
4 signals – 2 per minute	1,7
1 signal – 6 per minute	1,4
2 signals – 6 per minute	2
4 signals – 6 per minute	2,3,4

Descriptive statistics for signal response time are presented in Table 10 and Figure 6. ANOVA results are presented in Table 11. ANOVA results indicated a significant main effect for both condition and interaction of number of signals with signal frequency for signal response time ($F_{6, 343}$=18245, p≤0.001 and $F_{1, 343}$=2.890, p≤0.05, respectively). Post hoc tests showed significant differences for Condition 1 with all other conditions (p=0.0001), Condition 3 and 4 (p=0.008), and conditions 4 and 5 (p=0.048). Post hoc differences are listed in Table 12. Figure 6 depicts the response time data for all conditions except Condition 1 since the response time for that condition is 0.0. The plot for the six conditions where signals were presented is shown in Figure 7.

Table 10. Signal response time for model conditions (seconds)

Condition	Mean	Std. Dev.
No signals	0	0
1 signal – 2 per minute	0.689	0.002
2 signals – 2 per minute	0.693	0.002
4 signals – 2 per minute	0.686	0.002
1 signal – 6 per minute	0.691	0.002
2 signals – 6 per minute	0.689	0.002
4 signals – 6 per minute	0.691	0.002

Table 11. ANOVA for signal response time for model conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Condition	6	3.401	18253.2*
Signal Frequency	1	5.43×10^{-5}	0.292
Number of Signals	2	1.89×10^{-4}	1.016
Signal frequency * Number of signals	2	5.710×10^{-4}	3.070**
Error	343	1.86×10^{-4}	

* p≤0.001
** p≤0.05

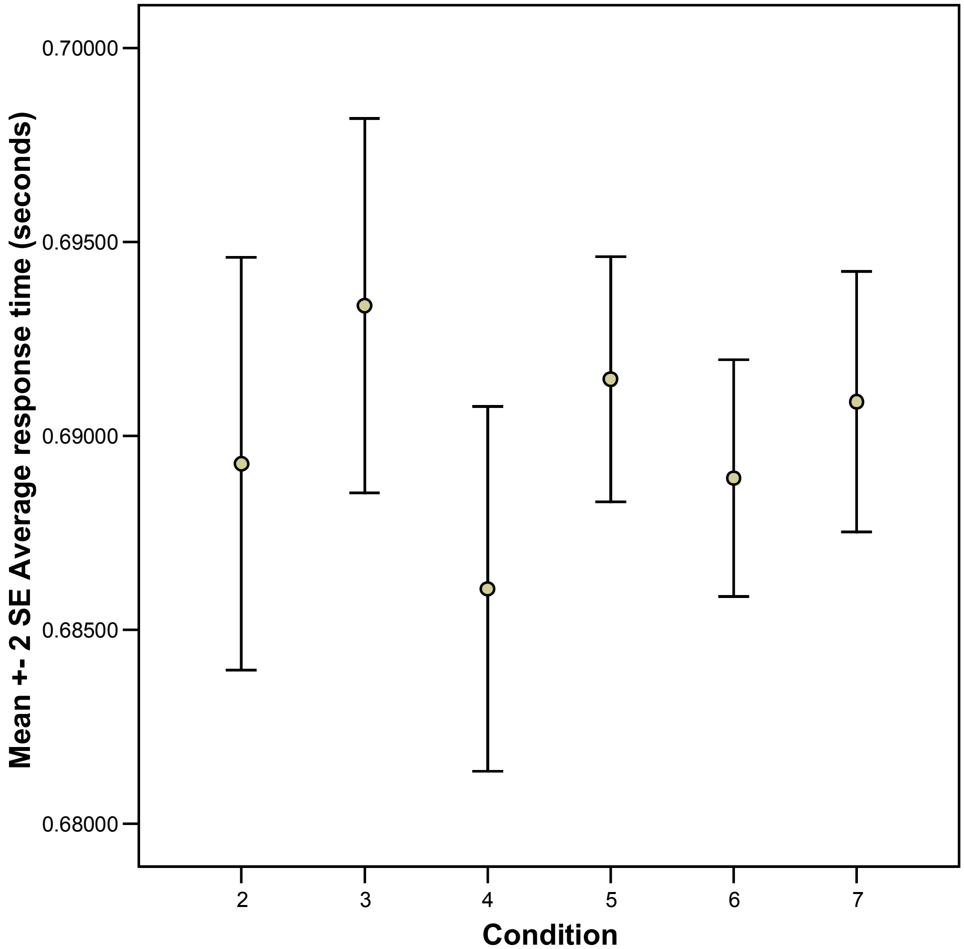

Figure 5. Signal response time for model conditions (seconds)

Table 12. Significant differences between conditions for signal response time in the model study

Condition	**Significantly Different Conditions**
No signals	2,3,4,5,6,7
1 signal – 2 per minute	1
2 signals – 2 per minute	1,4
4 signals – 2 per minute	1,3,5
1 signal – 6 per minute	1,4
2 signals – 6 per minute	1
4 signals – 6 per minute	1

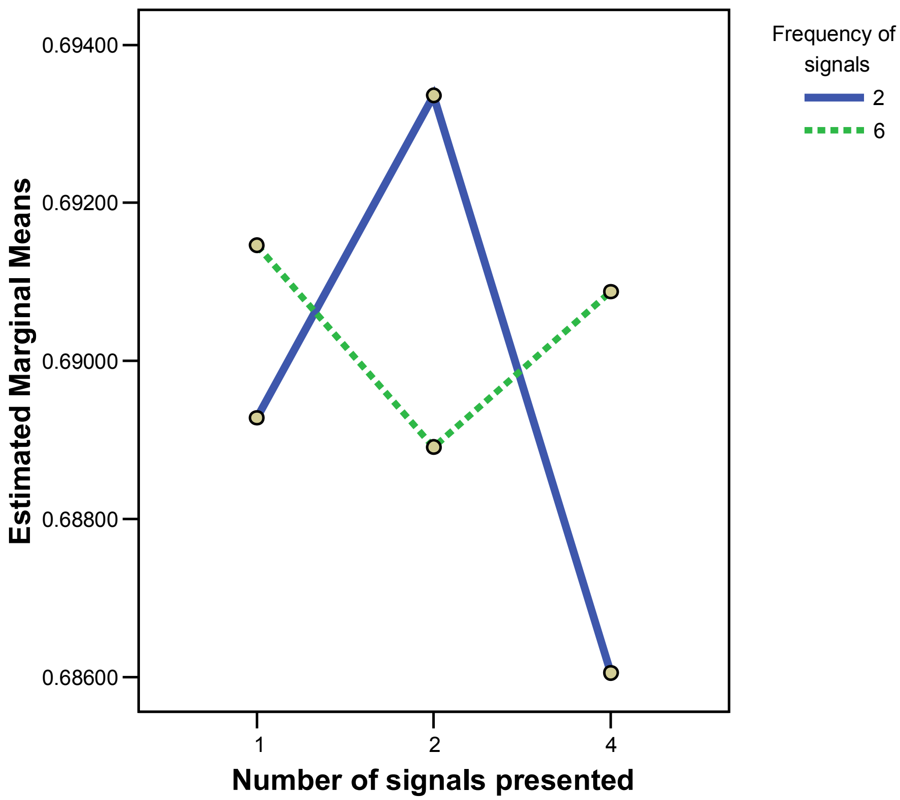

Figure 6. Signal response time interaction of number of signals and frequency of signals (seconds)

Descriptive statistics for percent time in overload are presented in Table 13 and Figure 9. ANOVA results are presented in Table 14. ANOVA results indicate a significant main effect for all factors; condition, frequency of signal, number of signals, and interaction of number of signals with signal frequency for percent time in overload ($F_{6,343}$=232.7, $p \leq 0.001$, $F_{6,343}$=563.0, $p \leq 0.001$, $F_{6,343}$=284.6, $p \leq 0.001$, and $F_{1,343}$=60.8, $p \leq 0.001$, respectively). Interaction is shown in Figure 9. Post hoc test show significant differences for Condition 1 and Conditions 4 through 7 (p=0.0001), Condition 2 and Conditions 4 through 7 (p=0.0001), Condition 3 and Conditions 4 through 7 (p=0.0001), Conditions 4 and 6 (p=0.012), Conditions 4 and 7 (p=0.0001), Conditions 5 and 6 (p=0.024), Conditions 5 and 7 (p=0.0001), and Conditions 6 and 7 (p=0.0001). Post hoc differences are listed in Table 15.

Table 13. Percent time in overload for model conditions

Condition	Mean	Std. Dev.
No signals	16.31	0.84
1 signal – 2 per minute	16.19	0.85
2 signals – 2 per minute	16.32	0.82
4 signals – 2 per minute	17.63	0.97
1 signal – 6 per minute	17.67	0.92
2 signals – 6 per minute	18.07	0.85
4 signals – 6 per minute	21.6	0.89

Table 14. ANOVA for percent time in overload for model conditions ($\alpha = 0.05$)

Source of Variance	Degrees of Freedom	Mean Square	F
Condition	6	179.246	232.662*
Signal Frequency	1	433.537	563.035*
Number of Signals	2	219.109	284.557*
Signal frequency * Number of signals	2	46.843	60.835*
Error	343	0.770	

* $p \leq 0.001$

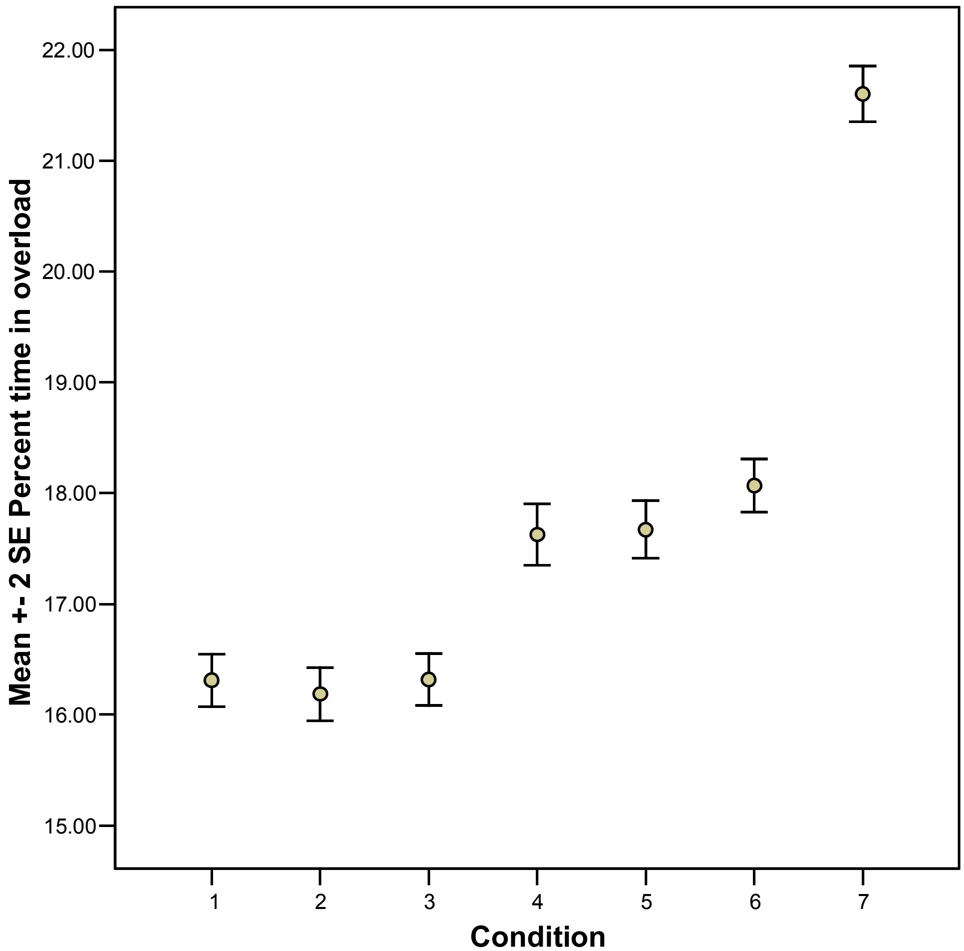

Figure 7. Percent time in overload for model conditions

Table 15. Significant differences between conditions for percent time in overload in the model study

Condition	**Significantly Different Conditions**
No signals	4,5,6,7
1 signal – 2 per minute	4,5,6,7
2 signals – 2 per minute	4,5,6,7
4 signals – 2 per minute	1,2,3,6,7
1 signal – 6 per minute	1,2,3,6,7
2 signals – 6 per minute	1,2,3,4,5,7
4 signals – 6 per minute	1,2,3,4,5.6

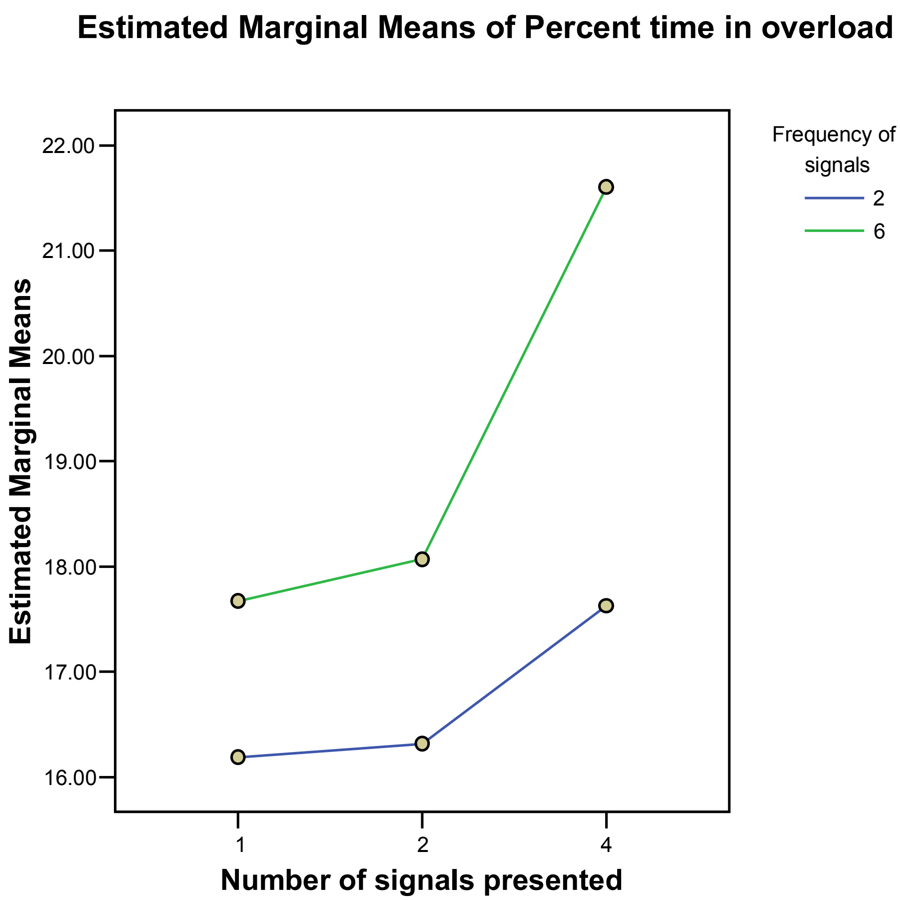

Figure 8. Interaction between number of signals and frequency of signals for percent time in overload for the model conditions.

4.2 Experiment 2 – Simulator Results

The simulator study included 14 human subjects. The demographics of the subjects were as follows. Seven participants were male and seven were female. The average age was 38.4 years old. All the participants were civilian HRED employees with the exception of one military participant. They had an average of 22 years of driving experience and spend an average of 37 hours per week working on a computer. One subject had extensive experience in video or computer games where a vehicle was controlled. The rest had none or minimal video or gaming experience (2 or less hours per month).

The results of the second experiment follow. The dependent measures that were collected included time on course, average speed, and response time for secondary task. Also, the NASA-TLX subjective workload scores were recorded for each trial.

The descriptive statistics for time to complete the course are shown in Tables 16. ANOVA results are shown in Table 17. For Experiment 2, ANOVA results indicated a significant difference for subject for time on course ($F_{6, 72}$=17.941, $p \leq 0.001$).

Table 16. Time to complete the course for study conditions (minutes)

Condition	Mean	Std. Dev.
No signals	5.08	0.56
1 signal – 2 per minute	5.16	0.35
2 signals – 2 per minute	5.14	0.56
4 signals – 2 per minute	5.19	0.57
1 signal – 6 per minute	5.17	0.53
2 signals – 6 per minute	5.14	0.51
4 signals – 6 per minute	5.09	0.61

Table 17. ANOVA for time to complete the course for study conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Subject	13	1.488	17.941*
Order	6	0.084	1.013
Condition	6	0.022	0.270
Signal Frequency	1	0.019	0.229
Number of Signals	2	0.006	0.072
Signal frequency * Number of signals	2	0.028	0.337
Error	72	0.083	

* $p \leq 0.001$

Descriptive statistics for average speed on the course are presented in Table 18. ANOVA results are presented in Table 19. ANOVA results indicated a significant main effect for subject for average speed on course ($F_{6, 72}$=18.024, $p \leq 0.001$).

Table 18. Average speed on the course for study conditions (miles per hour)

Condition	Mean	Std. Dev.
No signals	34.07	3.56
1 signal – 2 per minute	33.14	2.71
2 signals – 2 per minute	33.57	3.16
4 signals – 2 per minute	33.14	3.46
1 signal – 6 per minute	33.36	3.18
2 signals – 6 per minute	33.43	3.25
4 signals – 6 per minute	34	3.46

Table 19. ANOVA for average speed on the course for study conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Subject	13	55.834	18.024*
Order	6	3.592	1.159
Condition	6	1.997	0.645
Signal Frequency	1	2.012	0.649
Number of Signals	2	0.798	0.258
Signal frequency * Number of signals	2	1.798	0.580
Error	72	3.098	

* $p \leq 0.001$

Descriptive statistics for signal response time are presented in Table 20 and Figure 10. ANOVA results are presented in Table 21. ANOVA results indicated a significant main effect for subject, condition, and number of signals for signal response time ($F_{13, 72}$=9.280, $p \leq 0.001$, $F_{6, 72}$=91.161, $p \leq 0.001$, and $F_{2, 72}$=50.125, $p \leq 0.001$, respectively). Post hoc tests indicated significant differences for Condition 1 and all other Conditions (all p=0.000), Condition 2 and all other Conditions (p=0.0001(3), p=0.0001(4), p=0.025(5), p=0.001(6), p=0.0001(7)), Condition 3 and Conditions 4 and 5 (p=0.034(4) and p=0.0001(5)), Condition 4 and Conditions 5 and 6 (p=0.0001(5) and p=0.014(6)) and Condition 5 and Conditions 6 and 7 (both p=0.0001). These are listed in Table 22.

Table 20. Signal response time for study conditions (seconds)

Condition	Mean	Std. Dev.
No signals	0	0
1 signal – 2 per minute	0.63	0.19
2 signals – 2 per minute	0.82	0.19
4 signals – 2 per minute	0.92	0.24
1 signal – 6 per minute	0.52	0.13
2 signals – 6 per minute	0.8	0.25
4 signals – 6 per minute	0.89	0.19

Table 21. ANOVA for signal response time for study conditions ($\alpha = 0.05$)

Source of Variance	Degrees of Freedom	Mean Square	F
Subject	13	0.147	9.280*
Order	6	0.025	1.558
Condition	6	1.441	91.161*
Signal Frequency	1	0.058	3.625
Number of Signals	2	0.802	50.125*
Signal frequency * Number of signals	2	0.017	1.063
Error	72	0.016	

* $p \leq 0.001$

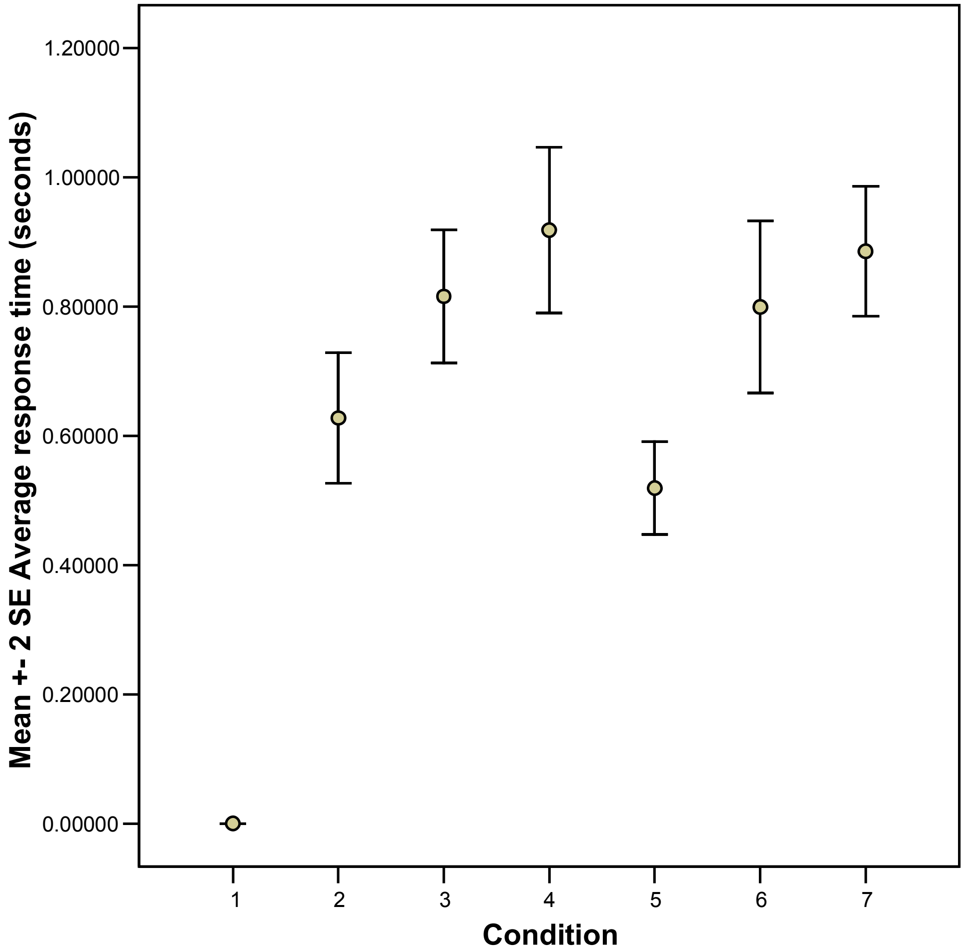

Figure 9. Signal response times for study conditions (seconds)

Table 22. Significant differences between conditions for signal response time in the simulator study

Condition	**Significantly Different Conditions**
No signals	2,3,4,5,6,7
1 signal – 2 per minute	1,3,4,5,6,7
2 signals – 2 per minute	1,2,4,5
4 signals – 2 per minute	1,2,3,5,6
1 signal – 6 per minute	1,2,3,4,6,7
2 signals – 6 per minute	1,2,4,5
4 signals – 6 per minute	1,2,5

Descriptive statistics for overall NASA-TLX scores are presented in Table 23 and Figure 11. ANOVA results are presented in Table 24. ANOVA results indicate a significant main effect for subject, condition, and frequency of signal for subjective workload scores ($F_{13, 72}$=31.809, p≤0.001, $F_{6, 72}$=2.458, p≤0.05, and $F_{1, 72}$=4.353, p≤0.05, respectively). Post hoc tests indicate a significant difference for condition 1 and conditions 4 through 7 (p=0.038(4), p=0.021(5), p=0.005(6), and p=0.002(7)) and conditions 2 and 7 (p=0.036). Post hoc differences are listed in Table 25.

Each NASA-TLX subscale was evaluated to determine if the independent factors had significant impact on any single component of workload. Only the mental demand subscale showed significant differences between conditions, but there was no significance for either independent factor; signal frequency or signal complexity. The ANOVA table for mental demand is shown in Table 26. No other subscale showed significant differences. The means of each subscale for all conditions are shown in Figure 12. This figure indicates that mental demand and effort were the highest contributors to workload, while physical demand and frustration were the lowest (Performance is the highest rated scale but higher performance in a lower contributor to workload.).

Table 23. NASA-TLX scores for study conditions

Condition	Mean	Std. Dev.
No signals	30.27	13.7
1 signal – 2 per minute	33.12	14.75
2 signals – 2 per minute	34.49	15.2
4 signals – 2 per minute	35.58	14.87
1 signal – 6 per minute	36.21	14.79
2 signals – 6 per minute	37.58	17.77
4 signals – 6 per minute	38.49	16.92

Table 24. ANOVA for NASA-TLX scores for study conditions (α = 0.05)

Source of Variance	Degrees of Freedom	Mean Square	F
Subject	13	1410.330	31.809*
Order	6	48.423	1.092
Condition	6	108.994	2.458**
Signal Frequency	1	193.021	4.353**
Number of Signals	2	39.604	0.893
Signal frequency * Number of signals	2	0.085	0.002
Error	72	44.338	

* p≤0.001
** p≤0.05

Table 25. Significant difference between conditions for NASA-TLX in the simulator study

Condition	Significantly Different Conditions
No signals	4,5,6,7
1 signal – 2 per minute	7
2 signals – 2 per minute	
4 signals – 2 per minute	1
1 signal – 6 per minute	1
2 signals – 6 per minute	1
4 signals – 6 per minute	1,2

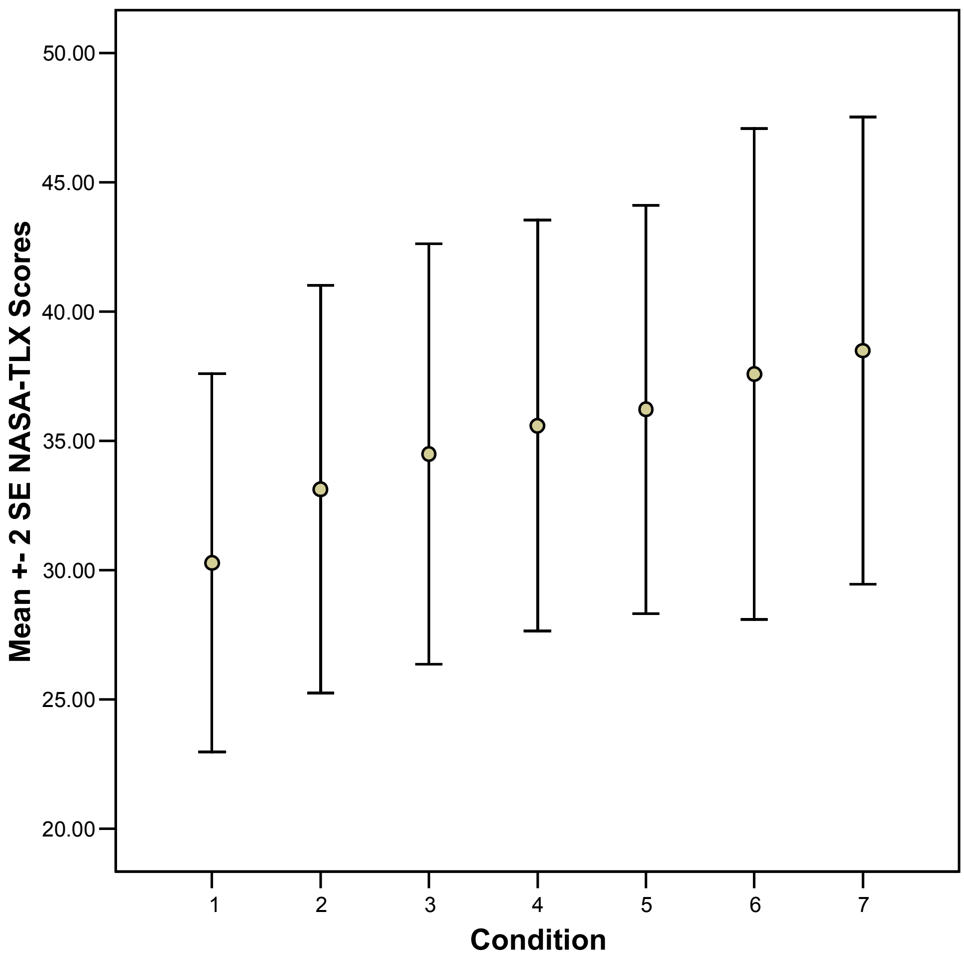

Figure 10. NASA-TLX scores for study conditions

Table 26. ANOVA table for mental demand subscale of NASA-TLX scores

Source of Variance	Degrees of Freedom	Mean Square	F
Subject	13	6503.046	54.092*
Order	6	182.088	1.515
Condition	6	327.279	2.722**
Signal Frequency	1	385.714	3.208
Number of Signals	2	258.143	2.147
Signal frequency * Number of signals	2	60.143	0.500
Error	72	120.223	

* p≤0.001

** p≤0.05

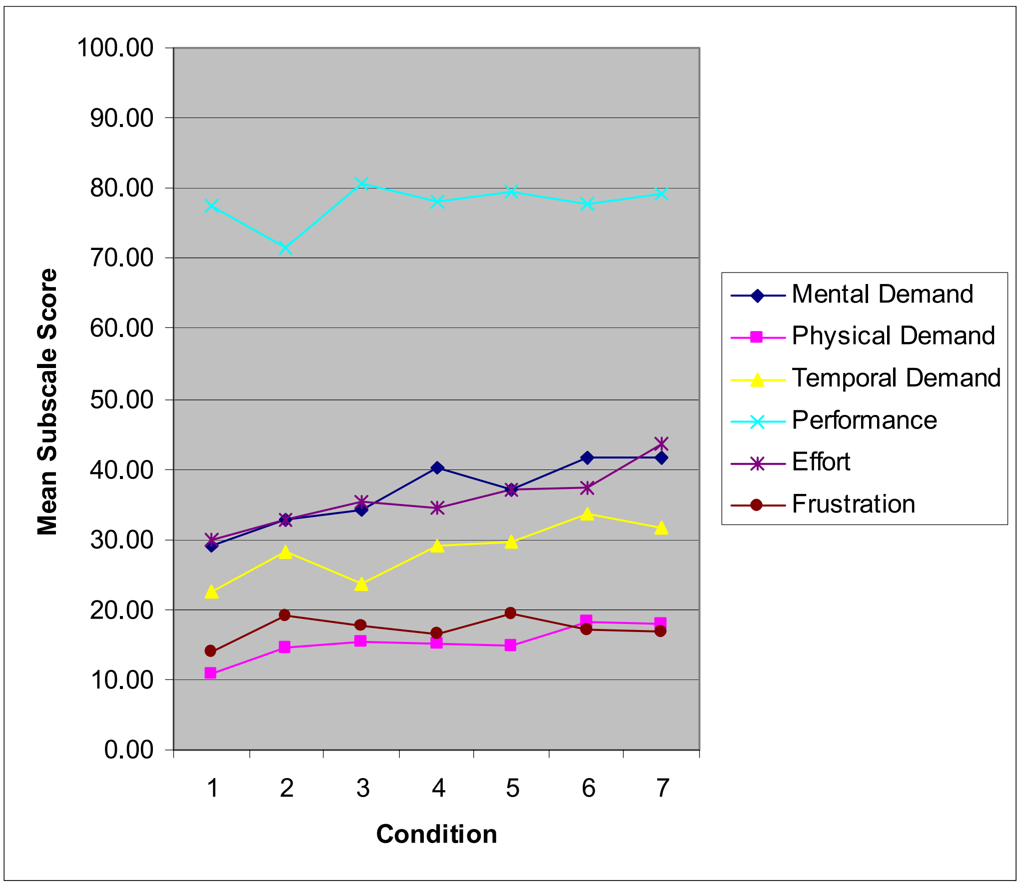

Figure 11. NASA-TLX Subscale scores for all study conditions

4.3 Comparison Analysis

Comparison of dependent measures from the model and simulator study was completed by correlation analysis. Correlation of all dependent measures is shown in Table 26. Negative correlations indicate that as one variable increases the other decreases. This would be expected for average speed and time on course. As the average speed increases the time on course would decrease. Positive correlations indicate a positive relationship between the variables. As one increases the other would also increase.

High correlations, those closer to 1.0 or -1.0 indicate that the relationship between the two variables is strong. Correlations are significant if their p-value is less than 0.05. This indicates that the relationship between the two variables would only happen by chance less than 5% of the time. This does not indicate a cause and effect relationship.

Correlations

		study - time to complete the course	study - average speed on the course	study - response time	study - NASA-TLX or % time in overload	model - time to complete the course	model - average speed on the course	model - response time	model - % time in overload
study - time to complete the course	Pearson Correlation	1	-.959**	.462	.203	.731	-.743	.633	-.318
	Sig. (2-tailed)	.	.001	.296	.662	.062	.056	.127	.487
	N	7	7	7	7	7	7	7	7
study - average speed on the course	Pearson Correlation	-.959**	1	-.441	-.141	-.784*	.791*	-.626	.376
	Sig. (2-tailed)	.001	.	.322	.763	.037	.034	.133	.406
	N	7	7	7	7	7	7	7	7
study - response time	Pearson Correlation	.462	-.441	1	.788*	.581	-.528	.895**	.459
	Sig. (2-tailed)	.296	.322	.	.035	.172	.223	.006	.300
	N	7	7	7	7	7	7	7	7
study - NASA-TLX or % time in overload	Pearson Correlation	.203	-.141	.788*	1	.022	.039	.764*	.777*
	Sig. (2-tailed)	.662	.763	.035	.	.963	.934	.045	.040
	N	7	7	7	7	7	7	7	7
model - time to complete the course	Pearson Correlation	.731	-.784*	.581	.022	1	-.997**	.619	-.350
	Sig. (2-tailed)	.062	.037	.172	.963	.	.000	.138	.442
	N	7	7	7	7	7	7	7	7
model - average speed on the course	Pearson Correlation	-.743	.791*	-.528	.039	-.997**	1	-.576	.419
	Sig. (2-tailed)	.056	.034	.223	.934	.000	.	.176	.349
	N	7	7	7	7	7	7	7	7
model - response time	Pearson Correlation	.633	-.626	.895**	.764*	.619	-.576	1	.321
	Sig. (2-tailed)	.127	.133	.006	.045	.138	.176	.	.483
	N	7	7	7	7	7	7	7	7
model - % time in overload	Pearson Correlation	-.318	.376	.459	.777*	-.350	.419	.321	1
	Sig. (2-tailed)	.487	.406	.300	.040	.442	.349	.483	.
	N	7	7	7	7	7	7	7	7

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 27. Correlations of dependent measures for model and simulator experiments.

Chapter 5 Discussion

5.1 Research Question 1 - Does auditory distraction impact the performance and workload of driving as depicted in the DWM?

To determine the answer to this question, the results of the experiments can be analyzed separately. The primary measure of interest in this investigation is workload since the purpose of the model is workload studies. Performance measures were collected also to determine if the workload level was impacting the performance. Performance measures will be discussed first and then workload.

5.1.1 Performance Measures –Time on course and average speed

5.1.1.1 Model results

In determining whether auditory secondary tasks impact driving performance and workload, one could start with the model results. Performance measures of time on course and average speed were impacted by the secondary task. Both time on course and average speed were different for the various conditions tested. Both dependent measures were significantly impacted by signal frequency ($p \leq 0.001$). Figures 4 and 5 show that time on course was longer and average speed slower for those conditions where the secondary task was presented less often.

5.1.1.2 Simulator results

Performance measures in the simulator study did not show significant differences. Lack of vestibular feedback could be a reason the simulator study did not exhibit significant difference in the performance measures. When the participant drove off the road because they took a corner too fast, it did not feel any different (i.e. no vestibular feedback) and there were no penalties. Therefore, participants were not anticipating corners and curves and slowing down. The drivers were instructed to "complete the course as quickly as possible while maintaining the vehicle on the path." It was observed that most participants placed more emphasis on speed than maintaining the vehicle on the road even though they were told that accuracy took precedence over time. It appeared that they tried to finish as fast as possible because they all went off the road on the sharp curves and made no apparent effort to anticipate the curves.

Also, the performance measures that were available on this particular simulator were not very sophisticated. Average speed is calculated by averaging the speed at each screen update. Time when the vehicle was stopped or moving in reverse (subjects would have to reverse if they struck an obstacle) were included in the average speed calculation. If the vehicle was stopped, speed would be zero and if in reverse, speed would be negative. The simulator collected the speed and time-on-course data for each screen update and averaged the sum of speeds by dividing by the number of updates. There were no means of evaluating specific portions of the course, for example, particular straight-aways or those portions less difficult to maneuver.

5.1.2 Performance Measure – Response time

5.1.2.1 Model results

An additional performance measure was the response time to the auditory signal or performance of the secondary task. The response times included in the model are derived from the micromodels built into the IMPRINT tool (Archer and Adkins, 1999). The micromodel used to represent the difference in response time is "choice reaction time" which is Hick's Law as described in Card, et al. (1983). The response time in the model showed significant differences between Condition 1 (no signals – baseline) and all other conditions. This is expected because Condition 1 did not have any secondary task therefore, no response. With the exception of conditions 3 (2 signals – 2 per minute) and 4 (4 signals – 2 per minute) and conditions 4 and 5 (1 signal – 2 per minute) the other conditions provide similar response times. The interaction of the two independent variables was significant as shown in Figure 7. This interaction is not expected as the response time should increase as the number of signals increased. This is an input into the model. The execution of the signal response tasks in the model should be examined for errors.

5.1.2.2 Simulator results

In the simulator study however, response time shows significant differences in many conditions. Response time increased as the number of response choices increased as shown in Figure 10. This is what was expected because choice reaction time increases with the number of choices.

5.1.3 Workload measures

5.1.3.1 Model results

The final dependent measure and the one of most interest was workload. In the model, workload was measured as the percent time the operator had an overall workload value greater than 40. Workload would be expected to be different for all conditions as workload is an input into the model. It is increased with the number of signals presented as shown in Figure 8. The probability of a workload value greater than 40 would occur more frequently with an increase in signal frequency as the model tasks that correspond to the signal response occur more frequently. The model showed significant differences in workload for the both number of signals and signal frequency. The interaction of signal frequency and complexity is also significant as shown in Figure 9. This would be expected because the change in workload value from 1 to 2 signals is much smaller than the change in workload value from 2 to 4 signals. Therefore the percent time in overload wouldn't change much with the increase in frequency at the complexity levels of 1 and 2 signals. However, when the workload value goes up much higher at the 4 signal complexity, the percent time in overload will be much greater at higher frequency.

5.1.3.2 Simulator results

In the simulator study, workload was recorded as subjective ratings of workload for the different conditions tested. ANOVA results indicated significant differences for condition and with frequency of signal as shown in Table 24. Workload changed with the increasing task demands as shown in Figure 11. This is what one would expect. However, there were no performance changes as a result of the increase in workload. This may be a result of the unsophisticated performance measures of the simulator or it may be a result of the dissociation of workload and performance as reported in Yeh and Wickens (1988). They report that subjective

workload assessments may show higher workload conditions than performance measures indicate when two tasks demand different resources.

Additionally, current research in brain activity shows that attentional shifts between visual attention and auditory attention required additional activity in the regions of the brain that normally attributed to the visual domain (Shomstein and Yantis, 2004). This also indicates that additional mental workload is required to process the auditory signals; therefore, the additional mental workload reported by NASA-TLX ratings is believed to be real. The NASA-TLX subscale analysis confirms that mental demand is the significant workload contributor across the conditions tested. Horrey and Wickens (2004) conducted a study on different display types for secondary tasks while driving. Auditory displays had more impact on certain aspects of driving performance than did visual displays. They attributed the difference to preemption of the primary task. Their secondary task was to repeat a string of numbers from four to ten digits long. Therefore the shift from visual to auditory was for a longer duration and required more working memory. This increase in duration and working memory is supported by the work conducted by Shomstein and Yantis (2004).

The impact of secondary task on performance is demonstrated in the model by the time-on-course and average speed measures. This is evidenced by the significance of signal frequency. The response time shows some of impact on performance in the study but this performance was on the secondary task and not the primary task. The impact of secondary task on workload was demonstrated in both the model and the simulator study.

5.2 Research Question 2 - Can IMPRINT correctly predict the impact on performance and workload?

5.2.1 Performance measures- Time on course and average speed

The two performance measures, average speed on course and time on course correlate between the model and the simulator study output. Table 27 shows the correlation matrix for these dependent measures for both studies. Correlation of average speed (model) with time on course for the model runs is significant at $p=0.0001$. Correlation of average speed (study) with time on course for the study is significant at $p=0.001$. This would be expected because average speed is a function of time on course. These correlations show a strong relationship between the average speed and time on course. The model correlation is higher as speed is a mathematical function of the time on course. In the study, the relationship is still very strong but because of the way speed is calculated (average speed for each screen update), the relationship is not a direct linear relationship.

Table 28. Correlation of performance measures for both experiments

Correlations

		model - time to complete the course	model - average speed on the course	study - time to complete the course	study - ave speed on course
model - time to complete the course	Pearson Correlation	1	-.997**	.731	-.784*
	Sig. (2-tailed)	.	.000	.062	.037
	N	7	7	7	7
model - average speed on the course	Pearson Correlation	-.997**	1	-.743	.791*
	Sig. (2-tailed)	.000	.	.056	.034
	N	7	7	7	7
study - time to complete the course	Pearson Correlation	.731	-.743	1	-.959**
	Sig. (2-tailed)	.062	.056	.	.001
	N	7	7	7	7
study - ave speed on course	Pearson Correlation	-.784*	.791*	-.959**	1
	Sig. (2-tailed)	.037	.034	.001	.
	N	7	7	7	7

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

However, the time on course for the study did not correlate with the average speed (model) or the time on course for the model runs. They were close to being significant at p=0.056 and p=0.062 respectively. This would indicate about a 6% chance of this relationship happening by chance. The differences may be due to the way the course was represented in the model. The decision to speed up or slow down in the model was probabilistically decided by the distance traveled at that point in the model or location on the course. The course was represented in the model by approximating a goal speed based on the location on the course. Location was determined by distance traveled. If a specific location on the course was situated in a sharp curve, the goal speed would be slower than the speed on a slight curve or in a straight section. There were no obstacles built into the model. Driving was represented as a set of continuous tasks that were broken into chunks. Basically a set of continuously repeating tasks represent the different aspects of driving. If the time chunk chosen was not small enough, the decisions to change speed may not have been frequent enough to adequately represent the simulator course. If the time chunks were too small, the data collection abilities of the model would be overcome. Trade-offs were made to best represent the continuous process and still be able to interpret the output data.

If replicating this particular course in the model were important, better calculations of location and goal speed would be required. This would hopefully allow better correlation of time on course and average speed between the model and the simulator. However, the actual use of this model is a generic representation of driving. Therefore, the specific course is not required to be exact.

The average speed on course for the study did correlate with both the time on course (model) and the average speed for the model runs (both p=0.03). This can be explained because the course distance in both cases is the same. The model was built with the assumption that the

vehicle would always be driven in the forward direction and would not stop. The actual representations of driving may differ but the changes from one condition to another were similar.

5.2.2 Performance measures – Response time

Even though the model response time differences between conditions are not significant, the correlation between the model response times and the study response times is high ($p=0.006$). This would indicate that the model may not predict exact performance results for the secondary task but performance trends are predicted in the model conditions. However, this correlation is questionable because the response time values from the model are questionable as stated in section 5.1.2.1.

5.2.3 Workload Measures

Performance on a secondary task can also be used as a measure of workload (Sanders and McCormick, 1993). If the performance of the primary task, in this case driving, does not change, changes in the performance of the secondary task would indicate differences in workload. In this study, the response time and NASA-TLX scores correlate ($p=0.035$). That gives two separate measures of workload for the conditions tested that correlate. This is further indication that the workload measured in the study is an increase in mental workload experienced by the driver.

Additionally, the workload measures in the model and the study correlate. The percent time in overload for the model runs correlates with the NASA-TLX scores ($p=0.040$). A positive correlation indicates that increases in workload from one condition to another in the simulator study are reflected by increases in mental workload in the model.

VACP workload measures primarily the mental demand imposed by the tasks being performed (psychomotor accounts for some physical). While NASA-TLX is designed to account for all aspects of workload including frustration and effort, the mental demand subscale was appeared to be the highest contributor to the overall scores. Both the VACP and the NASA-TLX subscale analysis indicate that driving is a "mental demand" task.

Predicting the workload and its impact on performance is the purpose of this model. The correlations of the results from the model and simulator study show that relative workload changes between conditions are predicted well in the model. The impact on performance of the changes in workload was not predicted well. This could be due to inadequate performance measures in the simulator or to poor representation in the model of the effects of increased workload on performance.

5.3 Research Question 3 – Is the DWM consistent with actual differences between less-demanding and more frequent distractions and fewer higher-demand distractions?

5.3.1 Model Results

The conditions with higher frequency of auditory signals had less of an impact on performance measures than did the less frequent signals evidenced by the increase in time on course and the decrease in average speed. Model results showed that frequency of signals had a significant impact on time on course and average speed on course. The number of signals presented did not impact the driving performance measures.

5.3.2 Simulator results

While the performance measures in the simulator study did not show significant differences, observations made during the data collection support the idea that less frequent auditory signals have more impact on performance. When the signals were displayed in 10 second blocks, the subjects were anticipating them and were not "startled" by the signal. However, in the conditions for which signals were presented in 30 second blocks, participants were not expecting them and reacted more visibly. They would momentarily jerk the steering wheel or release the accelerator for an instant.

Data from the model and observations from the simulator study support the conclusion that less frequent distractions have more of an impact than more frequent distractions. The data from the simulator did not support this conclusion. Additional studies would be necessary to determine if model coefficients need to be adjusted or if better performance measures are need for the simulator.

Chapter 6 Conclusions

The results of these experiments indicate that auditory distractions can impact workload experienced during driving. The same was not shown for performance. While the impact on performance was significant in the model, performance measures were not affected by the auditory tasks in the study. This may be partly due to the lack of sensitivity of the performance measures in the simulator and partly due to the dissociation of performance and workload caused by the alternate resources. However, there were performance differences in the secondary task that indicate there may be some impact on performance. Using the secondary task as an indicator of workload also supports the prediction that secondary tasks will impact workload. When a secondary task is added, workload is increased. Parameters in the model can be adjusted to reflect no performance changes at this level of workload increase following the model-test-model philosophy. The primary use of this model is to collect workload measures. Performance measures are usually not examined; however it is important to understand what impact of performance may result from the increases in workload.

The correlations between the model data and the study data indicate that the model may be valid in representing workload of driving. That is, the simulation predicts workload differences from condition to condition similar to those that were experienced in the simulator study. While this study validates workload in the model of operating this particular simulator, the model tasks are generic enough that this model can be used in other studies as representative of driving workload, particularly because this driving representation is used in comparison studies. The vehicle crew analyses where the DWM is included are not designed to look at driving specifically, but designed to investigate combat tasks that include driving. Driving tasks exhibit the same characteristics in all configurations. Therefore, overall workload of all combat tasks, including driving, is compared for the different configurations.

Secondary tasks will impact performance and workload, but the type of task will determine by how much. These investigations indicated that less frequent distractions had a greater impact on performance. Sheridan (2004) presented a control theory model of driver distraction. He stated that the impact of the distraction will depend on whether it occurs, "…at the state of driver intending, vehicle/environment state sensing, cognitive/action decision making, or vehicle response. (p. 598)" The overall conclusion based on the research cited and on this study is that driving is a task that requires considerable mental resources. Additional tasks give the potential for performance errors. Performance errors do not necessarily translate into accidents but they could. At a minimum, it should be assumed that additional risk is experienced when attempting secondary tasks while driving.

Most adults perform the common task of driving every day. Many perform additional tasks while driving, such as eating, talking on the cell phone, and grooming. While the majority of drivers do not have performance errors that lead to accidents daily, they may not realize how often they weave out of the lane, miss a stop light, or pull in front of someone. Driver distraction is of great interest currently. In 2004, a special section of *Human Factors* was devoted to driver distraction. In that issue, the preface states, "The (8) papers included in this special section demonstrate the diversity of potential distractions and diversity of methods to understand the safety consequences of these distractions. (p. 583)" (Lee and Strayer, 2004). Human performance modeling could be a useful means to better understand the impact of driving and distraction.

In the military environment, driving is even more difficult than the driving to which civilians are accustomed. It is often over terrain that is not maintained for vehicle travel. The terrain is usually unfamiliar to the driver and sometimes under hostile threat. It is difficult to convince people that driving requires full capacity and other tasks should be assigned elsewhere. This is partly due to the fact that most people perform other tasks while driving, even in a military environment. Driving for many people, in most situations, quickly becomes an automatic task and the performance errors that are committed are usually not critical and not always obvious.

Validation of the DWM is critical in supporting the argument that driving is a high workload task in our analysis of human performance in combat vehicles. This DWM is useful in presenting data that show the increase in workload when performing additional tasks while driving. For further validation of this model, one would need to show the potential for performance errors associated with the workload increase. Validation of the predicted workload of driving and the potential for performance impact is beneficial in military system design. Development of new designs for combat and non-combat vehicles will include some type of driving. Allocation of system functions between crew members and automation requires understanding of the mental workload imposed on the human operators.

This investigation shows that workload changes predicted by the model were demonstrated in the simulator. The performance changes predicted by the model were not seen in the simulator study. Further investigation would be required to assure the performance changes predicted by the model were real.

Chapter 7 Future Research

The simulator that was used in this study was limited in its ability to track performance measures. It would be interesting to complete the same study on a more sophisticated simulator that would allow measurement of lane deviations and speed changes when the secondary auditory task was presented.

Additionally, it is important to understand driving from a military point of view. Testing in actual military vehicles would increase the validity of the model data. In many of the new system concepts, drivers may be required to operate their vehicles with indirect vision. They would be driving from a computer screen, with views from outside shown through a camera. The simulator that was used in this study and many other simulators are better representations of indirect driving than actual driving. However, understanding the difference between simulators and direct or indirect driving is critical in system design (Van Erp and Padmos, 2003).

Additionally, mental workload is a challenging concept to measure. Determining a threshold for mental workload is difficult at best. One person may be able to complete two simultaneous tasks at a certain workload level while another may not. Also, mental workload is typically compared from one condition to another so that relative workload can be observed. Having a method to understand what a mental workload score means would be a great use for IMPRINT modeling.

If increases in mental workload could be compared to something such as blood alcohol content (BAC), analysts could better relate to the changes in mental workload for changing conditions. One could research studies where changes in taskings are related to BAC such as the Direct Line Motor Vehicle Study (2002) and determine what the change in mental workload would be. If the task changes could be related to BAC then perhaps mental workload could be related to BAC. This would have to be completed across several different studies to see if the same change in mental workload correlated with a similar change in BAC. This would provide a much clearer understanding of mental workload. People can more easily relate the possible impacts to performance if one could say for example, an increase in mental workload from 40 to 60 is similar to operating with a BAC of 0.10.

References

Archer, S. and Adkins, R. (1999). IMPRINT users guide. Draft prepared for US Army Research Laboratory.

Archer, S. and Allender, L. (2001). New capabilities in the army's human performance modeling tool. In *Proceedings of the Military, Government and Aerospace Simulation, 2001 Advanced Simulation Technologies Conference*, 22-27.

Army Research Laboratory Technical Advisory Board (ARL TAB). (2002). Private communication.

Balci, O. (1997). Verification, validation and accreditation of simulation models. In *Proceedings of the 1997 Winter Simulation Conference*, S. Andradŏttir, K.J. Healy, D.H. Withers, and B.L. Nelson, Eds. IEEE, Piscataway, NJ., 135-141.

Bierbaum, C.R., Szabo, S.M., and Aldrich, T.B. (1989). Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model. Volume I: Summary Report (AD-A210 763). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Biral, F. and Da Lio, M. (2001). Modelling Drivers with the Optimal Manoeuvre Method [On-line]. http://www.dinamoto.mecc.unipd.it/articoli%20moto/30articolo.pdf

Brown, T.L., Lee, J.D., and McGehee, D.V. (2000). "An attention-based model of driver performance in rear-end collision situations." *Transportation Research Record*, Vol. 1724, pp. 14-20.

Byers, J.C., Bittner, A.C., Jr., and Hill, S.G. (1989). Traditional and raw Task Load Index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety I*. London: Taylor & Francis.

Card, S. K., Moran, T.P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Cooke, N.J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies,* 41, pp. 801-849.

Defense Modeling and Simulation Office. (2005). Validation of Models and Simulation [On-Line]. http://vva.dmso.mil/Ref_Docs/Val_Lawref/Val-LawRef.htm.

Department of the Army. (1997). Management of Army Models and Simulations (AR 5-11). Washington, DC.

Department of the Army. (1999). Verification, Validation, and Accreditation of Army Models and Simulations (DA Pam 5-11). Washington, DC.

Direct Line Insurance (2002). The Mobile Phone Report [On-line]. http://info.directline.com/xxx/news.nsf/64125738690474fe00256a6f003a151b/bec9c738833c7fb180256b84002dec5f/$FILE/Mobile%20Phone%20Report.pdf.

Harrell, C. and Tumay, K. (1997). Simulation made easy. *Industrial Engineer*, (29)7; pp. 39-41.

Hart, S.G. and Staveland, L.E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human mental workload*. Amsterdam: North-Holland.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Jr., Zaklad, A.L., and Christ, R.E. (1992). Comparison of Four Subjective Workload Rating Scales. *Human Factors*, 34(4), pp. 429-439.

Horrey, W.J. and Wickens, C.D. (2004). Driving and Side Task Performance: The Effects of Display Clutter, Separation, and Modality. *Human Factors,* 46(4), pp. 611-624.

IMPRINT. (2004). http://www.arl.army.mil/ARL-Directorates/HRED/imb/imprint/Imprint7.htm. U.S. Army Research Laboratory.

Kirwan, B., and Ainsworth, L.K. (1992). *A guide to task analysis*. London: Taylor and Francis.

Law, A.M. and Kelton, W.D. (1991). *Simulation Modeling and Analysis.* New York, New York: McGraw-Hill, Inc.

Lee, J.D. and Strayer, D.L. (2004). Preface to the Special Section on Driver Distraction. *Human Factors,* 46(4), pp. 583-586.

Levison, W.H. (1993). A Simulation Model for Driver's Use of In-Vehicle Information Systems, *Transportation Research Record*, Vol. 1403. pp. 7-13.

McCracken, J. H. and T. B. Aldrich (1984). Analyses of selected LHX mission functions: implications for operator workload and system automation goals (Technical Note ASI479-024-84). Fort Rucker, AL: Army Research Institute Aviation Research and Development Activity.

McCormick, E.J. (1970). *Human Factors Engineering*. New York, New York: McGraw-Hill, Inc.

McMahon, R., Spencer, M. and Thorton, A. (1995). A Quick Response Approach to Assessing the Operational Performance of the AM93E1 NBCRS Through the Use of Modeling and Validation Testing. In *International Test and Evaluation Association Proceedings*. (Huntsville, AL). pp. 115-11.

Mitchell, D.K. (2000). Preliminary results of validation of HRED advanced workload modeling tools. In *Minutes of the 42nd Meeting of the Department of Defense Human Factors Engineering Technical Advisory Group,* Arlington, Virginia.

Mitchell, D.K. (2003). *Advanced Improved Performance Research Integration Tool (IMPRINT) Vetronics Technology Test Bed Model Development*. ARL-TN-0208. Aberdeen Proving Ground, MD: Army Research Laboratory.

Mitchell, D.K., Samms, C., Henthorn, MAJ T., and Wojciechowski, J.Q. (2003). Trade *Study: 2- vs. 3-soldier crew for the Mounted Combat System (MCS) and other Future Combat System platforms*. ARL-TR-3026. Aberdeen Proving Ground, MD: Army Research Laboratory.

Redelmeier, D.A. MD and Tibshirani, R.J. MD. (1997). Association Between Cellular-Telephone Calls and Motor Vehicle Collisions. *The New England Journal of Medicine*, 336(7), 453-458.

Robinson, S. (1997). Simulation Model Verification and Validation: Increasing the Users' Confidence [On-Line]. *http://www.informs-sim.org/wsc97papers/0053.PDF*

Rubio, S., Díaz, E., Martín, J., and Puente, J. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology: An International Review*, 53(1), pp.61-86.

Salvucci, D.D., Boer, E.R., and Lui, A. (2001). Toward an integrated model of driver behavior in cognitive architecture. *Transportation Research Record*, Vol. 1779, pp. 9-16.

Sanders, M.S. & McCormick, E.J. (1993). *Human Factors in Engineering and Design*. New York, NY: McGraw-Hill, Inc.

Schlegel, R.E. (1993). Driver Mental Workload. In Peacock and Karwowski (Eds.) *Automotive Ergonomics*. Washington D.C.: Taylor & Francis.

Shankle, R.R. (2003). Virtual Proving Ground Support FY2003. Final Report Contract DABJ05-03-F-002, U.S. Army Aberdeen Test Center.

Sheridan, T.B. (2004). Driver Distraction from a Control Theory Perspective. *Human Factors,* 46(4), pp. 587-599.

Shomstein, S. and Yantis, S. (2004). Control of Attention Shifts between Vision and Audition in Human Cortex. *Journal of Neuroscience,* 24(47), 10702-10706.

Slick, R.F., Cady, E.T., and Tran, T.Q. (2005). Workload Changes in Teenaged Drivers Driving with Distractions [On-Line]. http://www.drivesafety.com/03simulators/Workload_Changes_In_Teen_Drivers_with_Distractions.pdf

Strayer, D.L, Drews, F.A. and Johnston, W.A. (2003). Cell Phone Induced Failures of Visual Attention During Simulated Driving. *Journal of Experimental Psychology: Applied,* 9, 23-33.

Tijerina, L. (2000). Issues in the evaluation of driver distraction associated with in-vehicle information and telecommunications systems [On-line]. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/3.PDF

Van Erp, J.B.F. and Padmos, P. (2003). Image parameters for driving with indirect viewing systems. *Ergonomics,* 46(15), pp.1471-1499.

Wickens, C. and Hollands, J. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall.

Wierwille W.W. (1993). Visual and manual demands of in-car controls and displays. In Peacock and Karwowski (Eds.) *Automotive Ergonomics*. Washington D.C.: Taylor & Francis.

Wojciechowski, J.Q. and Archer, S. (2002). Human Performance Modeling: Art or Science? *Proceedings of the 2002 Military, Government and Aerospace Simulation Symposium: Advanced Simulation Technologies Conference*. San Diego, CA. pp. 59-64.

Wojciechowski, J.Q., Kogler, T.M., and Lockett, J.F. Jr. (2001). Examining Direct, Teleoperated, and Semi-Autonomous Off-Road Driving From a Human Information Processing Perspective. Unpublished manuscript.

Yeh, Y. and Wickens, C.D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors,* 30(1), 111-120.

Appendix A.

VACP Workload Scales from IMPRINT (IMPRINT, 2004)

Scale Value	Visual Scale Descriptor
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.7	Visually Discriminate (detect visual differences)
4.0	Visually Inspect/Check (discrete inspection/static condition)
5.0	Visually Locate/Align (selective orientation)
5.4	Visually Track/Follow (maintain orientation)
5.9	Visually Read (symbol)
7.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)

Scale Value	Auditory Scale Descriptor
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound)
2.0	Orient to Sound (general orientation/attention)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound)
4.9	Interpret Semantic Content (speech)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)

Scale Value	Cognitive Scale Descriptor
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (consider single aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion

Scale Value	Psychomotor Scale Descriptor
0.0	No Psychomotor Activity
1.0	Speech
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight control, sensor control)
4.6	Manipulative
5.8	Discrete Adjustive (rotary, vertical thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)

Appendix B.

Driver Workload Model Diagrams

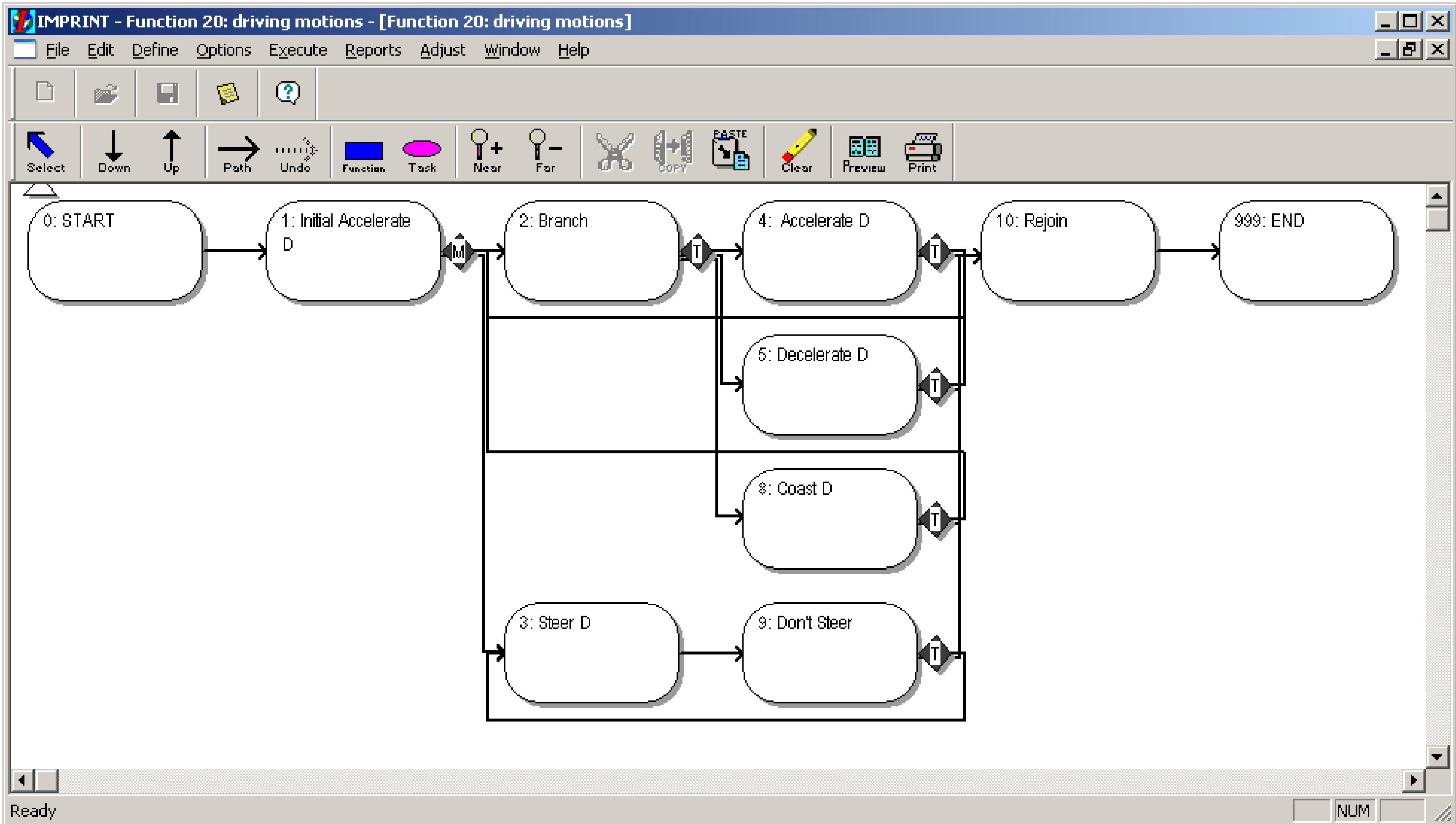

Figure 12. Move function from the DWM

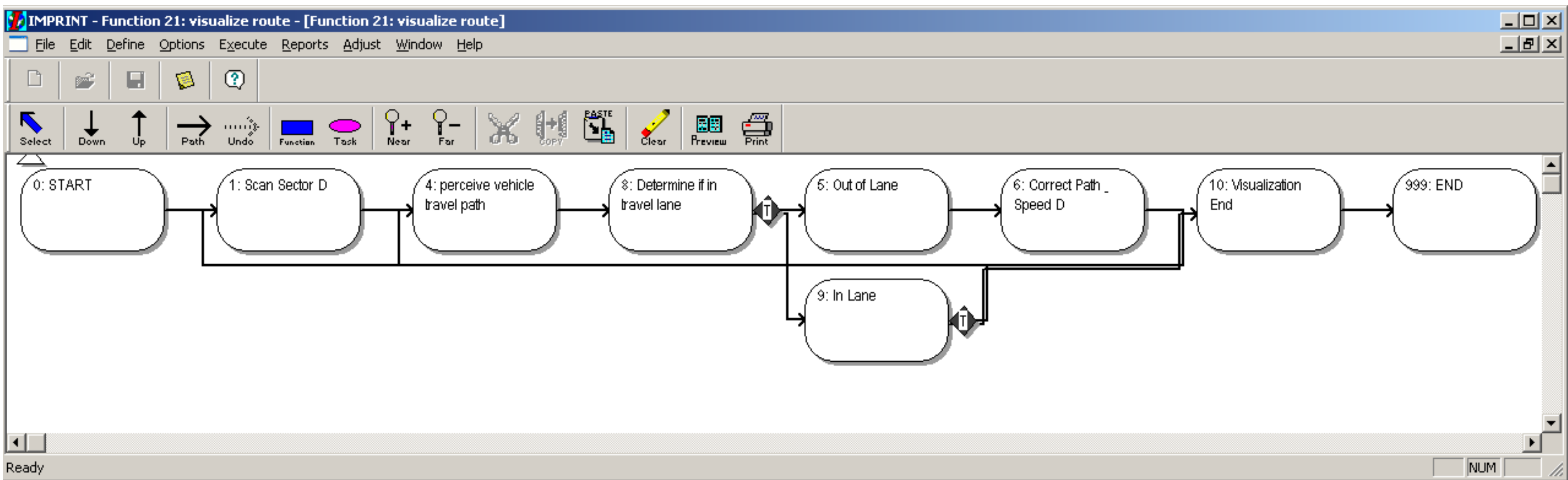

Figure 13. See function from the DWM.

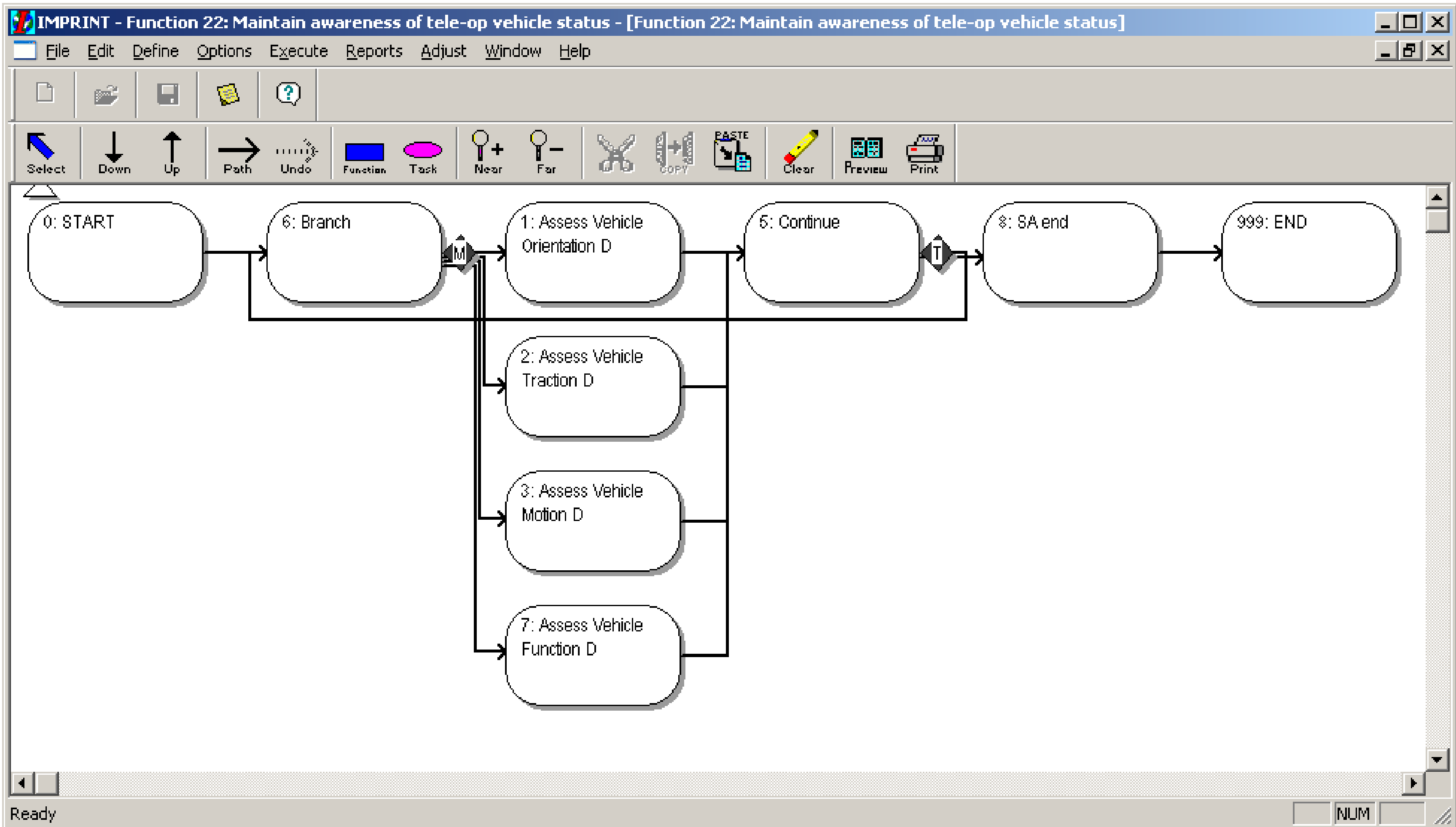

Figure 14. Maintain Situation Awareness function from the DWM.

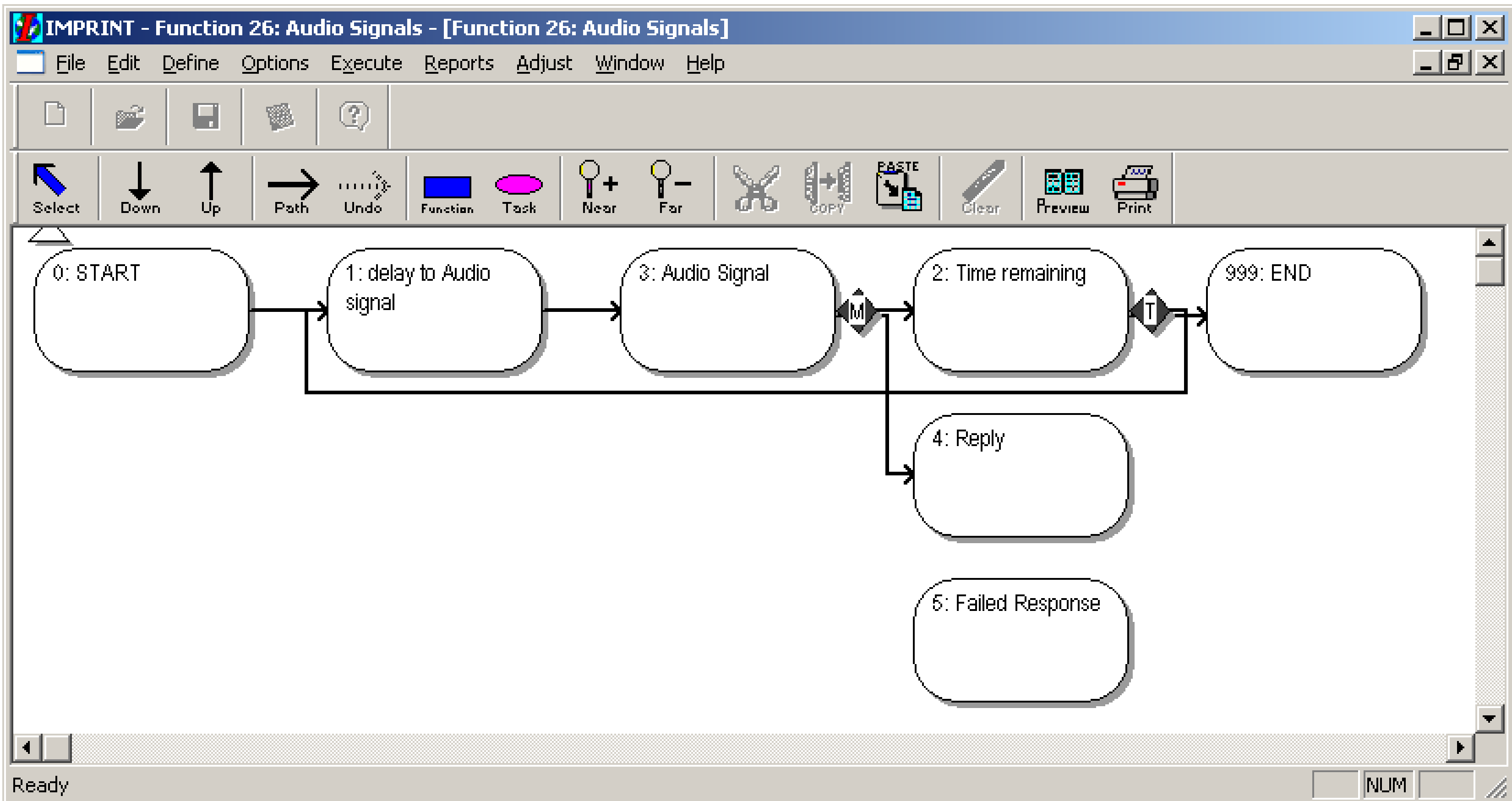

Figure 15. Signal-Response function from the DWM.

Appendix C.

Questionnaires

Participant ID:________________Date:__________________Condition:___________________________

Demographics and Personal Experience Form (read to participants and completed by experimenter)

1. Age: _________2. Gender: ________

3. Medical Data:
a. Which is your dominant hand? _______________

b. Do you wear glasses when working on the computer? YesNo

c. Have you ever experienced Moderate to severe Motion Sickness? Yes No Simulator Sickness?
Severity?_________________ Other Comments ____
____

4. Educational Data

a. What is your highest level of education received?
____GED ____High School ____Some College ____Bachelors Degree ____M.S/M.A
____Ph.D. Other______________________
b. What subject is your degree in (if applicable?, example Engineering) __________________

5. Military Data (for current military personnel):

a. Grade:	E1	E2	E3	E4	E5	E6	E7	E8	E9
	O1	O2	O3	O4	O5				
	WO1	CWO2	CWO3	CWO4	CWO5				

b. Primary MOS/AFSC: _________
c. Time in MOS/AFSC Years:_______ Months______
d. Duty Position/Title: ____
e. Time in present duty position: Years: ______ Months: ______
f. Length of service? Years: ______ Months: ______

7. Prior Experience

a. Do you know how to drive a car or truck? (circle one) Yes No
b. How many years of experience do you have in driving a car or truck? _____________years
c. How many hours per week do you use a computer at home or at work? ______________hours
d. Do you have any experience with computer games where you control a vehicle? (for example, driving simulators, race car simulators, aircraft simulators) (circle one) Yes No
--If Yes, what kind of computer game? _____________________________
--How much time do you spend playing this game? ______________ hours per month
--Have you ever felt funny (e.g., dizzy, queasy, disoriented) after playing computer games?_____

Participant
ID:________________Date:__________________Condition:___________________________

TLX Workload Scale

Please rate your workload by putting a mark on each of the six scales at the point which matches your experience.

Mental Demand

Very Low Very High

Physical Demand

Very Low Very High

Temporal Demand

Very Low Very High

Performance

Very Low Very High

Effort

Very Low Very High

Frustration

Very Low Very High

Vita

Josephine Q. Wojciechowski was born in Brunswick, GA, USA. She received a Bachelor of Chemical Engineering from Georgia Institute of Technology in 1982. She spent 4 years working in industry at EI Dupont De Nemours and Bailey Controls Company. She went to work for the U.S. Army Ballistic Research Laboratory in 1986. In 1993, she transferred to the Human Research and Engineering Directorate of the U.S. Army Research Laboratory. She has spent over 10 years building human performance models of soldier tasks. She completed her Master of Science in Industrial and Systems Engineering, Human Factors Option in 2006.

NO. OF COPIES	ORGANIZATION
1 (PDF ONLY)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	US ARMY RSRCH DEV & ENGRG CMD SYSTEMS OF SYSTEMS INTEGRATION AMSRD SS T 6000 6TH ST STE 100 FORT BELVOIR VA 22060-5608
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC IMS 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
2	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS OK T 2800 POWDER MILL RD ADELPHI MD 20783-1197
	ABERDEEN PROVING GROUND
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK (TECH LIB) BLDG 4600
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK TP S FOPPIANO BLDG 459
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL HR BLDG 459